大学数学科学丛书 12

矩阵不等式

（第二版）

王松桂　吴密霞　贾忠贞　编著

科学出版社

北京

内 容 简 介

本书系统地论述了矩阵论中的各种不等式. 全书共分九章. 第 1 章是矩阵论的预备知识; 第 2~8 章分别讨论了有关秩、行列式、特征值、条件数、迹、偏序和受控等方面的不等式; 第 9 章给出了矩阵不等式在线性统计中的几个应用; 最后两个附录收集了数量、函数和概率统计中常用的不等式.

本书读者对象为高等院校高年级本科生、研究生、有关专业的教师与数学工作者及工程技术人员.

图书在版编目(CIP)数据

矩阵不等式/王松桂, 吴密霞, 贾忠贞编著. —2 版. —北京: 科学出版社, 2006.5

(大学数学科学丛书; 12)

ISBN 7-03-016494-6

I. 矩… II. ①王… ②吴… ③贾… III. 矩阵-不等式 IV. O151.25

中国版本图书馆 CIP 数据核字(2005)第 137711 号

责任编辑: 吕 虹 赵彦超 / 责任校对: 赵桂芬

责任印制: 安春生 / 封面设计: 王 浩

科学出版社 出版

北京东黄城根北街16号

邮政编码: 100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2006 年 5 月第 二 版 开本: B5(720×1000)

2006 年 5 月第一次印刷 印张: 18 1/4

印数: 1—4 000 字数: 336 000

定价: 39.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

《大学数学科学丛书》编委会

《大学数学科学丛书》序

按照恩格斯的说法，数学是研究现实世界中数量关系和空间形式的科学．从恩格斯那时到现在，尽管数学的内涵已经大大拓展了，人们对现实世界中的数量关系和空间形式的认识和理解已今非昔比，数学科学已构成包括纯粹数学及应用数学内含的众多分支学科和许多新兴交叉学科的庞大的科学体系，但恩格斯的这一说法仍然是对数学的一个中肯而又相对来说易于为公众了解和接受的概括，科学地反映了数学这一学科的内涵．正由于忽略了物质的具体形态和属性、纯粹从数量关系和空间形式的角度来研究现实世界，数学表现出高度抽象性和应用广泛性的特点，具有特殊的公共基础地位，其重要性得到普遍的认同．

整个数学的发展史是和人类物质文明和精神文明的发展史交融在一起的．作为一种先进的文化，数学不仅在人类文明的进程中一直起着积极的推动作用，而且是人类文明的一个重要的支柱．数学教育对于启迪心智、增进素质、提高全人类文明程度的必要性和重要性已得到空前普遍的重视．数学教育本质是一种素质教育；学习数学，不仅要学到许多重要的数学概念、方法和结论，更要着重领会到数学的精神实质和思想方法．在大学学习高等数学的阶段，更应该自觉地去意识并努力体现这一点．

作为面向大学本科生和研究生以及有关教师的教材，教学参考书或课外读物的系列，本丛书将努力贯彻加强基础、面向前沿、突出思想、关注应用和方便阅读的原则，力求为各专业的大学本科生或研究生（包括硕士生及博士生）走近数学科学、理解数学科学以及应用数学科学提供必要的指引和有力的帮助，并欢迎其中相当一些能被广大学校选用为教材，相信并希望在各方面的支持及帮助下，本丛书将会愈出愈好．

李大潜

2003 年 12 月 27 日

第二版前言

本书是矩阵不等式方面的第一部中文专著. 1994 年第一版出版以后, 深受广大读者的欢迎, 并远销海外, 成为数学工作者、高等院校有关专业教师和研究生乃至一般科技工作者手中的重要参考书或工具书.

为了更好地满足广大读者的需求, 在科学出版社的支持下, 我们对原书进行了修订. 这次修订除更正了第一版的一些错误外, 主要增加了近年来许多重要的新结果. 主要包括：Wielandt 不等式的矩阵形式 (7.7 节)；凸函数的矩阵不等式 (7.8 节)；Hadamard 乘积 (7.9 节)；Log- 弱受控不等式 (8.7 节) 以及一些关于矩阵行列式、迹和矩阵幂的迹的不等式 (分别添加在 3.2 节、6.3 节和 6.6 节适当部分), 总计增加了三十多个不等式. 这些增补内容的初稿, 由吴密霞博士完成, 最后由王松桂定稿.

尽管我们在进行修订时, 尽了最大的努力, 限于水平, 不当乃至谬误之处, 在所难免, 诚请广大读者指正.

作　者

2005 年 7 月

第一版序

正如作者在本书前言中所说, 撰写一部系统地、全面地论述矩阵论中各种不等式的专著, 是他们的一个夙愿. 经过十多个春秋的努力, 这个夙愿终于成了现实. 我作为他们的一名同事和同行, 特别是目睹了他们为创作这一专著而辛苦耕耘的过程, 感到由衷的高兴.

“十年辛苦不寻常”, 这确不是一部泛泛之作. 首先, 作者之一王松桂教授自 20 世纪 70 年代后期起即从事线性统计的研究, 多年来在国内外刊物上发表了一系列的论文和专著. 矩阵不等式是进行这种研究的一个基本工具. 所以, 本书的写作有他多年研究工作的素养和经验作为背景. 其次, 作者 10 年来访问了欧美各国许多著名的学术研究中心, 与国际上线性统计领域内卓有成就的同行进行了广泛的合作和切磋, 收集了许多最新的、国内不易见到的材料, 这些都大大提高了本书的学术价值和实用价值.

本书收集的材料很多, 这是一个显著的优点. 因为在通常的矩阵论教本和专著中, 对矩阵不等式这个题材多未作系统介绍, 许多结果散布在大量文献中, 使用和查找不便. 如今有了本书, 一卷在手, 可免除许多查找翻检之劳. 可贵的是, 内容虽多, 但作者努力做到了多而不乱, 多而不偏. 众多的材料按其性质组织成一些专题, 重点突出, 系统性强. 作者注意了所选材料在应用上的意义, 把它作为取舍的一项重要标准, 相信凡是细读过本书的人, 都会有这种感觉.

作者之一王松桂教授是一位统计学家. 因此可以理解, 作者在材料选择上重视其在统计学上的应用, 这是本书的一个特色. 这个特色, 加上材料收集之丰富, 使本书必将成为线性统计方面的工作者案头必备的工具, 也是统计和相近专业的研究生、大学生有用的参考读物. 这当然不是说本书的读者仅限于这些人. 相反, 由于矩阵这个工具在数学、自然科学、工程技术乃至社会科学中的重要性, 众多的读者将会发现, 他们都能从本书中找到一些对自己工作有用的东西. 我想, 这也是作者自己的愿望.

陈希孺

第一版前言

关于不等式, 已经出版了若干部英文专著, 其中最有影响的是 Hardy, Littlewood 和 Polya(1934, 1953) 的 “Inequalities”, Beckenback 和 Bellman(1961) 的 “Inequalities” 以及 Marshall 和 Olkin(1979) 的 “Inequalities: Theory of Majorization and its Applications”. 这些书或以数量和函数的不等式为主要讨论对象, 或从某一特定方面研究一类数量或矩阵的不等式. 随着矩阵理论的迅速发展及其在自然科学、工程技术和社会经济等领域的广泛应用, 关于矩阵不等式的新结果层出不穷, 它们或是经典不等式的改进和推广, 或是完全新型的不等式, 或是应用的深入或拓广. 这些结果都散见在各种刊物或著作之中, 对理论研究者和使用矩阵工具的广大科技工作者带来诸多不便. 多年来, 作者有一个夙愿, 就是广泛收集、整理各种涉及矩阵的不等式, 撰写一部系统、全面地论述矩阵论中的各种不等式的专著. 目前在国内外尚未见有这样内容的著作出版.

自 1987 年以来, 我们就开始在教学实践、科学研究和广泛的国内外学术交流中收集资料, 特别在跟欧洲朋友们的学术磋商中, 获益匪浅. 随着工作的深入, 我们发现有关矩阵不等式的文献之多出乎预料, 于是我们不得不在资料筛选上花费很多精力, 同时我们又特别注意一些最新结果 (本书文献截止至 1992 年 10 月), 力求使本书尽可能反映这一方向的全貌.

全书共分九章, 矩阵论中的各种不等式按秩、行列式、特征值、条件数、迹、偏序和受控等内容分类, 每一类构成一章. 有些结果就其内容讲, 既可排在这一章, 也可排在另一章. 对这样的内容, 为读者查阅方便, 我们把它们排在两处, 但只在其中一处给出证明. 为便于读者使用并使本书内容大体上做到自成体系, 在第 1 章系统叙述了矩阵论的预备知识, 多数结果给出了证明. 为了显示矩阵不等式在线性统计中的广泛应用, 第 9 章举了几个例子. 限于本书的性质和篇幅, 当然举例不可能是全面的. 最后, 附录 1 和附录 2 罗列了常见的数量、函数不等式以及概率统计中的重要不等式, 以便于读者参考.

在这里我们要特别感谢很多的朋友. 他们是芬兰 Tampere 大学 Liski 博士和 Puntanen 博士, 瑞典 Umea 大学 Kulldorff 教授, 瑞士 Berne 大学 Riedwyl 教授, 波兰 Poznan 大学 Baksalary 教授, 英国 Manchester 大学 Farebrother 教授, 加拿大 Ottawa 大学邵军教授和 J. N. K. Rao 教授, 加拿大 Waterloo 大学吴建福教授, 美国 Chicago 大学刁锦寰教授以及 Colorado 州立大学 Srivastava 教授等. 本书许多资料是我们于 1988~1989 和 1990~1991 年先后访问这些地方时收集的, 这些朋友的热情帮助和支持对本书的写作来说是非常重要的. 这里还要提到第一作者的女儿

王晓京和王晓天, 感谢她们在美国读书期间帮助我们复印了一些国内不易找到的文献. 作者还要感谢严利清和杨亚宁同志为本书出版所提供的帮助. 我们还要借此机会, 向安徽教育出版社杨晓原同志表示谢意, 感谢他为本书的出版所给予的自始至终的热情支持、大力帮助和有效合作.

最后, 在本书出版之际, 作者还要对我们的老师陈希孺教授多年来的谆谆教诲、指导、关心和帮助表示诚挚的谢意.

本书前三章由贾忠贞执笔, 后六章由王松桂执笔, 最后由王松桂定稿.

限于作者的水平, 本书不妥乃至谬误之处在所难免, 恳请国内同行和广大读者不吝赐教.

王松桂　贾忠贞

符 号 表

符号	含义
A'	矩阵 A 的转置
$\bar{A}$	矩阵 A 的共轭
A^*	矩阵 A 的共轭转置 (即 $\bar{A}'$)
A^{-1}	矩阵 A 的逆矩阵
A^-	矩阵 A 的广义逆矩阵
A^+	矩阵 A 的 Moore-Penrose 广义逆
$A \geqslant 0$	表示 A 为半正定阵 (实对称阵或 Hermite 阵)
$A > 0$	表示 A 为正定阵 (实对称阵或 Hermite 阵)
$A \otimes B$	A 与 B 的 Kronecker 乘积
$A \circ B$	A 与 B 的 Hadamard 乘积
$\|A\|$	矩阵 A 的任一范数
$\|A\|_F$	矩阵 A 的欧氏范数, 即 Frobenius 范数
$\|A\|_2$	矩阵 A 的谱范数
$A^{(k)}$	矩阵 A 的 k 阶复合阵
$A^{1/2}$	半正定阵 A 的半正定平方根
$\boldsymbol{C}^n$	所有 n 维复向量的全体
$\det A$	方阵 A 的行列式
D	$\{x \in R^n : x_1 \geqslant \cdots \geqslant x_n\}$
$\mathrm{Im}(z)$	z 的虚部
$k(A)$	矩阵 A 的条件数
$\mathscr{M}(A)$	矩阵 A 的列向量张成的子空间
$r(A)$	矩阵 A 的秩
R^n	所有 n 维实向量的全体
R^n_+	$\{x \in R^n : x_i \geqslant 0, i = 1, \cdots, n\}$
$\mathrm{Re}(z)$	z 的实部
$\mathrm{tr} A$	矩阵 A 的迹
$x < y$	向量 x 受控于 y
$x <_\omega y$	向量 x 弱受控于 y
$\|X\|$	向量 X 的欧氏长度
$\bar{z}$	z 的共轭复数
$\lvert z \rvert$	z 的模
$\lambda(A)$	方阵 A 的特征值
$\sigma(A)$	矩阵 A 的奇异值

目　　录

*　　　*　　　*

第 1 章　矩阵论的预备知识

本书的目的是系统地论述有关矩阵的各种不等式. 因此, 当写作本书时, 假定读者已经具备了一般线性代数教科书中矩阵论的知识. 但是, 为了读者阅读上的方便和叙述简洁, 在这一章, 将扼要地给出本书讨论中要用到的一些重要结论和一般文献中不易查到的事实. 因为本章前六节内容偏重于基础, 为了节省篇幅, 略去了部分证明, 读者可从许以超 (1965)、蒋尔雄 (1978) 等找到它们的证明. 而对其余各节, 所有结论都给出了较详细的证明.

§1.1　线 性 空 间

我们用 C^n 和 R^n 分别表示全体 n 维复向量和实向量组成的线性空间.

设 S 为 C^n 的一个子空间, $a_1,\cdots,a_k$ 为 S 的一组基. 记 $A=(a_1,\cdots,a_k)$ 是一个 $n\times k$ 矩阵. 则 S 可以表为

$$S=\{x\colon\ x=At, t\in C^k\},$$

即 S 是由 $a_1,\cdots,a_k$ 的所有线性组合生成的线性子空间, 称为 A 的列向量张成的子空间, 简称为 A 的列空间, 记为 $\mathscr{M}(A)$. 若用 $\dim(S)$ 表示 S 的维数, 则从矩阵秩和空间维数的定义可以推出

$$\dim\mathscr{M}(A)=r(A), \tag{1.1.1}$$

这里 $r(A)$ 表示 A 的秩. 注意, S 是零维子空间当且仅当 $S=\{0\}$, 即单个零向量构成的子空间.

设 S_1 和 S_2 为两个子空间, 那么它们的和

$$S_1+S_2=\{x+y\colon\ x\in S_1, y\in S_2\}$$

以及它们的交

$$S_1\cap S_2=\{x\colon\ x\in S_1\ 且\ x\in S_2\}$$

都是子空间, 分别称为 S_1 与 S_2 的和空间、交空间. 设 A 和 B 为两个具有相同行数的矩阵, 则容易证明

$$\mathscr{M}(A\vdots B)=\mathscr{M}(A)+\mathscr{M}(B). \tag{1.1.2}$$

下面的定理给出了和空间与交空间维数之间的关系.

定理 1.1.1 设 S_1 和 S_2 为两个子空间, 则

$$\dim(S_1+S_2)=\dim(S_1)+\dim(S_2)-\dim(S_1\cap S_2). \tag{1.1.3}$$

证明 记 $p_i=\dim S_i,\ i=1,\ 2.\ r=\dim(S_1\cap S_2)$, 设 $a_1,\cdots,a_r$ 为 $S_1\cap S_2$ 的基, 将 $a_1,\cdots,a_r$ 扩充为 $a_1,\cdots,a_r,b_1,\cdots,b_{p_1-r}$, 使其构成 S_1 的一组基. 同样地, 将 $a_1,\cdots,a_r$ 扩充为 $a_1,\cdots,a_r,c_1,\cdots,c_{p_2-r}$, 使其为 S_2 的一组基. 那么, 向量组

$$a_1,\cdots,a_r,b_1,\cdots,b_{p_1-r},c_1,\cdots,c_{p_2-r}$$

构成了 S_1+S_2 的基, 这就证明了结论.

特别地, 若 $S_1\cap S_2=\{0\}$, 则称 S_1+S_2 为 S_1 与 S_2 的直和, 记为 $S_1\oplus S_2$, 直和具有下列性质.

定理 1.1.2 (1) $\dim(S_1\oplus S_2)=\dim S_1+\dim S_2$;

(2) S_1+S_2 为直和 $\Longleftrightarrow$ S_1+S_2 中的任一向量能唯一地表成 S_1 与 S_2 中向量之和.

对于 C^n 中的任意两个向量 a 和 b, 它们的内积 (a,b) 定义为

$$(a,b)=b^*a,$$

这里 b^* 表示向量 b 的转置共轭向量. 当 $a,b\in R^n$ 时, $(a,b)=b'a$. 定义了内积的线性空间 R^n 和 C^n 分别称为欧氏空间和酉空间.

在内积空间中, 向量 a 的长度定义为

$$\|a\|=(a,a)^{1/2}.$$

当 $(a,b)=0$ 时, 称 a 与 b 正交, 记为 $a\perp b$. 一组向量 $a_1,\cdots,a_m$, 若

$$(a_i,a_j)=\begin{cases}1, & i=j,\\ 0, & i\neq j,\end{cases}$$

则称 $a_1,\cdots,a_m$ 为标准正交向量组. 如果对一切 $b\in S$, 总有 $a\perp b$, 则称 a 与子空间 S 正交, 记为 $a\perp S$. 根据子空间的定义, 容易验证, 向量集合

$$S^{\perp}=\{x\colon x\perp S\}$$

也是一个子空间, 称为 S 的正交补空间.

正交补空间具有下列性质.

定理 1.1.3 (1) $S\oplus S^{\perp}=C^n$;

(2) $S=(S^{\perp})^{\perp}$;

(3) $S_1 \subset S_2 \Longleftrightarrow S_2^{\perp} \subset S_1^{\perp}$;

(4) $(S_1 + S_2)^{\perp} = S_1^{\perp} \cap S_2^{\perp}$;

(5) $(S_1 \cap S_2)^{\perp} = S_1^{\perp} + S_2^{\perp}$.

对于一个 $m \times k$ 矩阵 A, 若 $m \times t$ 矩阵 B 的秩为 t, 且满足 $\mathscr{M}(B) = \mathscr{M}(A)^{\perp}$, 则以后把 B 记为 $A^{\perp}$. 于是 $\mathscr{M}(A^{\perp}) = \mathscr{M}(A)^{\perp}$.

在本节一开头, 我们定义了矩阵 A 的列空间 $\mathscr{M}(A)$, 现在引进与 A 相联系的另一个子空间. 众所周知, 齐次线性方程组 $Ax = 0$ 的解的全体构成一个子空间, 我们称它为 A 的零空间或矩阵 A 的核, 记为 $N(A)$, 即

$$N(A) = \{x\colon Ax = 0\}.$$

若用 A^{-} 表示 $A_{n\times k}$ 的广义逆，即满足条件 $AXA = A$ 的任一矩阵 X(详见§1.7), 则有

$$N(A) = \mathscr{M}(I_k - A^{-}A), \tag{1.1.4}$$

这里 I_k 表示 k 阶单位阵.

因为 $I - A^{-}A$ 和 $A^{-}A$ 都是幂等阵, 利用幂等阵的迹等于它的秩, 我们有 $r(I_k - A^{-}A) = k - \mathrm{tr}(A^{-}A) = k - r(A)$, 这里 $\mathrm{tr}(\cdot)$ 表示方阵的迹, 也就是对角线元素之和, 结合 (1.1.1), 我们有下面的结论.

定理 1.1.4 对任意 $n \times k$ 矩阵 A,

$$\dim \mathscr{M}(A) + \dim N(A) = k.$$

§1.2 特征值与特征向量

特征值与特征向量是矩阵论中两个重要的基本概念, 无论在理论研究或工程技术领域, 都有着广泛的应用.

定义 1.2.1 设 A 为 n 阶方阵, 若数 λ 和非零向量 x 满足

$$Ax = \lambda x, \tag{1.2.1}$$

则称 λ 为 A 的特征值, x 为 A 的对应于 λ 的特征向量.

线性方程组 (1.2.1) 可改写为 $(\lambda I_n - A)x = 0$. 此方程组有非零解当且仅当系数行列式 $\det(\lambda I_n - A) = 0$. 因此, λ 为 A 的特征值当且仅当 $\det(\lambda I_n - A) = 0$. 记 $\varphi(\lambda) = \det(\lambda I_n - A)$, 它是 λ 的 n 次多项式, 称为 A 的特征多项式. 于是, λ 为 A 的特征值当且仅当它是特征多项式 $\varphi(\lambda)$ 的根. 正是这个原因, 特征值也常常称为特征根.

对于特征多项式, 我们有下面的重要结论.

定理 1.2.1　设 n 阶方阵 $A=(a_{ij})$ 的特征多项式展开为

$$\varphi(\lambda)=\det(\lambda I_n-A)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_{n-1}\lambda+c_n, \tag{1.2.2}$$

则它的系数

$$c_k=(-1)^k\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\det \mathrm{A}(i_1,\cdots,i_k),\quad k=1,\cdots,n,$$

这里 $A(i_1,\cdots,i_k)$ 表示由 A 的第 $i_1,\cdots,i_k$ 行、列元素组成的 k 阶子阵, 称为 A 的主子阵. 而和号 $\sum\limits_{1\leqslant i_1<\cdots<i_k\leqslant n}$ 表示对所有可能的 1 至 n 中的整数 $i_1,\cdots,i_k$ 求和, 特别

$$c_1=-\mathrm{tr}A, \tag{1.2.3}$$

$$c_n=-\det A, \tag{1.2.4}$$

这里 $\mathrm{tr}A=\sum\limits_{i=1}^n a_{ii}$, 称为 A 的迹.

记 $\lambda_1(A),\cdots,\lambda_n(A)$ 为 A 的 n 个特征值, 则

$$\varphi(\lambda)=\prod_{k=1}^n(\lambda-\lambda_k(A)). \tag{1.2.5}$$

比较 (1.2.2) 和 (1.2.5), 结合 (1.2.3) 和 (1.2.4), 我们得到

推论 1.2.1　对任意的 n 阶方阵 A, 有

$$\mathrm{tr}A=\sum_{k=1}^n\lambda_k(A),$$

$$\det A=\prod_{k=1}^n\lambda_k(A).$$

利用特征值的定义和特征多项式, 容易证明特征值的下述性质.

定理 1.2.2　设 A 为 n 阶方阵, 则

(1) A' 和 A 有相同的 (包括重数) 特征值;

(2) A^* 的特征值是 A 的特征值的共轭, 即

$$\lambda(A^*)=\bar{\lambda}(A),$$

这里 $\lambda(A)$ 表示 A 的特征值, 而 $\bar{\lambda}(A)$ 表示共轭复数;

(3) 若 A 为可逆阵, 则

$$\lambda(A^{-1})=\frac{1}{\lambda(A)}.$$

进一步, 若 A 的特征值全为实数, 且排列为 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$, 则

$$\lambda_k(A^{-1}) = \frac{1}{\lambda_{n-k+1}(A)}, \quad k = 1, 2, \cdots, n.$$

在特征值的讨论中, 矩阵相似是一个重要的概念.

定义 1.2.2 设 A 和 B 为 $n \times n$ 阵. 若存在可逆阵 P, 使得

$$B = P^{-1}AP,$$

则称 B 和 A 相似, 记为 $B \sim A$.

容易验证, 矩阵相似具有下列性质:

(1) 自反性: $A \sim A$;

(2) 对称性: 若 $A \sim B$, 则 $B \sim A$;

(3) 传递性: 若 $A \sim B$, $B \sim C$, 则 $A \sim C$.

设 $B = P^{-1}AP$, 并且 $\varphi_B(\lambda)$ 表示 B 的特征多项式. 则

$$\begin{aligned}\varphi_B(\lambda) &= \det(\lambda I_n - P^{-1}AP) = \det(P^{-1}(\lambda I_n - A)P) \\ &= \det P^{-1} \cdot \det(\lambda I_n - A) \cdot \det P \\ &= \det P^{-1}P \cdot \det(\lambda I_n - A) \\ &= \det(\lambda I_n - A) = \varphi_A(\lambda).\end{aligned}$$

于是, 相似矩阵 B 与 A 有相同的特征多项式, 所以它们有相同的特征值.

定理 1.2.3 相似矩阵有相同的特征值.

推论 1.2.2 相似矩阵有相同的迹和行列式.

从理论的角度讲, 如果一个方阵能相似于对角阵, 则很容易求到它的特征值与特征向量. 设 A 为 n 阶方阵, 若 $\Phi = (\varphi_1, \cdots, \varphi_n)$ 为可逆阵, 使得

$$\Phi^{-1}A\Phi = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

等价地

$$A\Phi = \Phi \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

也就是

$$A\varphi_k = \lambda_k \varphi_k, \quad k = 1, \cdots, n.$$

于是 $\lambda_1,\cdots,\lambda_n$ 就是 A 的特征值, $\varphi_1,\cdots,\varphi_n$ 就是对应的特征向量, 并且这些特征向量线性无关. 所以, 如果一个方阵 A 能够相似于一个对角阵, 则该对角阵的对角元就是 A 的特征值, A 相似于对角阵时所用的可逆阵的列向量就是 A 的 n 个线性无关的特征向量. 反过来, 若 A 有 n 个线性无关的特征向量, 则以它们为列向量的矩阵, 就能使 A 相似于对角阵. 于是, 我们有

定理 1.2.4　n 阶方阵 A 相似于对角阵, 当且仅当 A 有 n 个线性无关的特征向量.

因为属于不同特征值的特征向量总是线性无关的, 所以

推论 1.2.3　若方阵 A 的特征值互不相同, 则 A 一定相似于对角阵.

众所周知, 并不是每个方阵一定能够相似于对角阵. 能够相似于对角阵的最重要的矩阵类是实对称阵和 Hermite 阵, 这将在下面两节详细讨论, 这里我们讨论一般情况.

设 λ 为 n 阶方阵 A 的一个特征值, 任何满足 $Ax=\lambda x$ 的非零向量 x, 都是 A 的对应于 λ 的特征向量. 易见, A 的对应于 λ 的全体特征向量是线性方程组 $(\lambda I_n-A)x=0$ 的全部非零解. 从线性方程组解的理论知, 如果添上零向量, 这些解就构成一个线性子空间, 记为 S_λ, 称为 A 的对应于特征值 λ 的特征子空间, 根据上节引进的零空间的定义, 有

$$S_\lambda=N(\lambda I_n-A).$$

特征子空间 S_λ 的维数称为特征值 λ 的几何重数, 因此几何重数就是对应于 λ 的所有特征向量中线性无关的最大个数. 如果一个 n 阶方阵 A, 它的所有特征值的几何重数之和为 n, 则它一定可以相似于对角阵, 遗憾的是, 这个假设并不总是成立的.

若 λ 为 A 的 m 重特征值, 也就是说, 它是特征多项式的 m 重根, 则称 m 为 A 的代数重数. 关于代数重数和几何重数, 我们有如下定理.

定理 1.2.5　对任一特征值, 几何重数总不超过代数重数.

证明　设 A 为 n 阶方阵, λ_0 为其任一特征值. 记它的几何重数为 g, 代数重数为 m, 则存在 g 个对应于 λ_0 的线性无关的特征向量 $\varphi_1,\varphi_2,\cdots,\varphi_g$, 将其扩充为 C^n 的一组基

$$\varphi_1,\cdots,\varphi_q,\varphi_{g+1},\cdots,\varphi_n. \tag{1.2.6}$$

记 $\Phi=(\varphi_1,\cdots,\varphi_g,\varphi_{g+1},\cdots,\varphi_n)$, 则

$$A\Phi=(\lambda_0\varphi_1,\cdots,\lambda_0\varphi_g,A\varphi_{g+1},\cdots,A\varphi_n).$$

因为向量组 (1.2.6) 是 C^n 的基. 于是 $A\varphi_{g+1},\cdots,A\varphi_n$ 均可由它们线性表出.

设

$$A\varphi_j=\sum_{k=1}^n b_{kj}\varphi_k,\quad j=g+1,\cdots,n,$$

并记

$$B=\begin{pmatrix}\lambda_0 I_g & B_1\\ 0 & B_2\end{pmatrix},$$

这里 $B_1=(b_{jk}),\ j=1,\cdots,g;k=g+1,\cdots,n$. 而 $B_2=(b_{jk}),j=g+1,\cdots,n;k=g+1,\cdots,n$, 则有 $A\Phi=\Phi B$, 也就是 $A\sim B$. 根据定理 1.2.3 前面的证明, A 与 B 有相同的特征多项式, 但 B 的特征多项式

$$\varphi_B(\lambda)=\det(\lambda I_n-B)=(\lambda-\lambda_0)^g\det(\lambda I_{n-g}-B_2).$$

于是, λ_0 的代数重数 m 至少是 g, 这就证明了 $g\leqslant m$. 证毕.

因为一个 n 阶方阵的所有特征值的代数重数之和总是等于 n, 所以综合定理 1.2.5 和前面的讨论, 我们证明了下面的定理.

定理 1.2.6 方阵 A 相似于对角阵, 当且仅当它的每一个特征值的几何重数总是等于它的代数重数.

在后面的讨论中, 我们常常要碰到两个矩阵乘积的特征值问题, 关于这一方面, 我们有下面的重要定理.

定理 1.2.7 (1) 设 A 和 B 分别为 $m\times n$ 和 $n\times m$ 阵, 则 AB 与 BA 有相同的非零特征值 (包括重数);

(2) 设 A、B 为 n 阶方阵, 且至少一个可逆, 则 AB 与 BA 相似.

证明 (1) 设 $\lambda\neq 0$, 则

$$\begin{pmatrix}I_m & -A\\ 0 & I_n\end{pmatrix}\begin{pmatrix}\lambda I_m & A\\ B & I_n\end{pmatrix}=\begin{pmatrix}\lambda I_m-AB & 0\\ B & I_n\end{pmatrix},$$

$$\begin{pmatrix}\lambda I_m & A\\ B & I_n\end{pmatrix}\begin{pmatrix}I_m & -\frac{1}{\lambda}A\\ 0 & I_n\end{pmatrix}=\begin{pmatrix}\lambda I_m & 0\\ B & I_n-\frac{1}{\lambda}BA\end{pmatrix}.$$

对上面两式两边分别取行列式, 得到

$$\det(\lambda I_m-AB)=\lambda^m\det\left(I_n-\frac{1}{\lambda}BA\right)=\lambda^{m-n}\det(\lambda I_n-BA). \tag{1.2.7}$$

得证

(2) 是下面事实的直接结果.

$$AB=\begin{cases}A(BA)A^{-1}, \text{若}A\text{可逆},\\ B^{-1}(BA)B, \text{若}B\text{可逆}.\end{cases}$$

在 (1.2.7) 中, 命 $\lambda=1$, 便得到下面的事实.

推论 1.2.4 设 A 和 B 分别为 $m\times n$ 和 $n\times m$ 矩阵, 则

$$\det(I_m-AB)=\det(I_n-BA).$$

§1.3　实对称阵

设 $A=(a_{ij})$ 的元素 a_{ij} 全是实数, 且 $A'=A$, 则称 A 为实对称阵. 实对称阵具有许多重要性质.

定理 1.3.1　设 A 为 n 阶实对称阵. 则

(1) A 的所有特征值都是实数;

(2) 存在正交阵 Φ, 使得

$$\Phi' A \Phi = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \tag{1.3.1}$$

即实对称阵一定正交相似于对角阵.

注 1　从 (1.3.1) 知

$$A \sim \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

所以, 由上节的讨论知, $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值.

记 $\Phi = (\varphi_1, \cdots, \varphi_n)$, 从 (1.3.1) 可推出

$$A\Phi = \Phi \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

也就是 $A\varphi_j = \lambda_j \varphi_j$, 所以 $\varphi_1, \cdots, \varphi_n$ 为 A 的 n 个标准正交化特征向量.

注 2　记 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 则 (1.3.1) 可改写为

$$A = \Phi \Lambda \Phi' = \sum_{j=1}^{n} \lambda_j \varphi_j \varphi_j', \tag{1.3.2}$$

此式称为 A 的谱分解.

从 (1.3.1), 我们很容易得到下面的推论.

推论 1.3.1　设 A 为 n 阶实对称阵, x 为 $n \times 1$ 实向量, 则存在正交阵 Φ, 使得正交变换 $y = \Phi' x$ 把二次型 $x'Ax$ 化为平方和

$$x'Ax = y'\Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \tag{1.3.3}$$

这里 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

在实对称阵中, 半正定阵是一个特别重要的矩阵类.

设 A 为 n 阶实对称阵, 若对任意 $x \in R^n, x'Ax \geqslant 0$, 则称 A 为半正定阵, 记为 $A \geqslant 0$. 若进一步, $x'Ax = 0$ 当且仅当 $x = 0$, 则称 A 为正定阵. 记为 $A > 0$. 因此在以后的讨论中, 若无特殊声明, 当我们使用记号 $A \geqslant 0$ 或 $A > 0$ 时, 总假定 A 是对称的. 下面的两个定理分别表征了半正定阵和正定阵, 它们是非常有用的.

定理 1.3.2　设 A 为 n 阶实对称阵, 则 $A \geqslant 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值都是非负的;

(2) 存在对称阵 B, 使得 $A = B^2$;

(3) 存在 $t \times n$ 矩阵 B, 这里 $t = r(A)$, 使得 $A = B'B$;

(4) 对任意矩阵 P, $P'AP \geqslant 0$;

(5) A 的所有主子式为非负, 这里主子式定义为主子阵 $A(i_1, \cdots, i_k)$ 的行列式, 也就是 A 的位于第 $i_1, \cdots, i_k$ 行、列交叉处元素组成的 k 阶方阵的行列式.

注 3 对半正定阵 A, 满足 (2) 的矩阵 B 是很多的. 下面构造出一个半正定的 B. 设 $A \geqslant 0, \lambda_1, \cdots, \lambda_n$ 为其特征值. $\varphi_1, \cdots, \varphi_n$ 为其对应的特征向量, 且 $\Phi = (\varphi_1, \cdots, \varphi_n)$ 为正交阵, 则由定理 1.3.1 之 (2). 有

$$
\begin{aligned}
A &= \Phi \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \Phi' \end{pmatrix} \Phi' \\
&= \Phi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \Phi' \Phi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \Phi' \\
&= B^2,
\end{aligned}
$$

其中 $B = \Phi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \Phi' \geqslant 0.$

以后我们把 B 记为 $A^{1/2}$, 称为 A 的平方根.

定理 1.3.3 设 A 为 n 阶实对称阵, 则 $A > 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值都是正数;

(2) 存在可逆对称阵 B, 使得 $A = B^2$;

(3) 存在可逆阵 B, 使得 $A = B'B$;

(4) 对任意可逆阵 $P, P'AP > 0$;

(5) A 的所有主子式为正数;

(6) A 的所有顺序主子式为正数, 这里 k 阶顺序主子式为主子阵 $A(1, 2, \cdots, k)$ 的行列式.

在本节的剩余部分, 我们讨论两个实对称阵同时对角化的问题.

定理 1.3.4 设 A, B 为两个 n 阶实对称阵, 且 $B > 0$. 则存在可逆阵 Q, 使得

$$A = Q'\Lambda Q, \tag{1.3.4}$$

$$B = Q'Q, \tag{1.3.5}$$

这里 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \lambda_j (j = 1, \cdots, n)$ 为 AB^{-1} 的特征值.

证明　因 $B > 0$, 故 $B^{-1} > 0$, 记 $B^{-1/2} = (B^{-1})^{1/2} > 0$. 因为 $B^{-1/2}AB^{-1/2}$ 仍为实对称阵, 依定理 1.3.1, 存在正交阵 Φ, 使得 $\Phi'(B^{-1/2}AB^{-1/2})\Phi = \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 记 $Q = \Phi' B^{1/2}$, 即得 (1.3.4) 和 (1.3.5), 这里 $\lambda_1, \cdots, \lambda_n$ 为 $B^{-1/2}AB^{-1/2}$ 的特征值. 根据定理 1.2.7(2), $B^{-1/2}AB^{-1/2}$ 与 AB^{-1} 相似, 所以, 它们有相同的特征值, 定理得证.

为了证明下面一个重要定理, 我们需要如下引理.

引理 1.3.1　设 A 为 n 阶实对称阵, x 为任一 n 维非零向量. 则一定存在 A 的特征向量 φ, 使得

$$\varphi \in \mathscr{M}(x, Ax, A^2x, \cdots). \tag{1.3.6}$$

证明　因为向量组 $\{x, Ax, A^2x, \cdots\}$ 总是线性相关的, 故一定存在 k, 使得

$$A^k x + c_{k-1}A^{k-1}x + \cdots + c_0 x = 0. \tag{1.3.7}$$

假设 (1.3.7) 中的 k 是具有这一性质的最小 k.

记

$$f(y) = y^k + c_{k-1}y^{k-1} + \cdots + c_0.$$

依代数学基本定理, $f(y) = 0$ 有 k 个根 $\mu_1, \cdots, \mu_k$. 于是 $f(y)$ 可分解为

$$f(y) = (y - \mu_1)(y - \mu_2)\cdots(y - \mu_k).$$

用 $f(A)$ 表示将上式中 y 换上 A, μ_i 换上 $\mu_i I_n$ 所得到的矩阵, 则 (1.3.7) 变形为

$$\begin{aligned} f(A)x &= (A - \mu_1 I)(A - \mu_2 I)\cdots(A - \mu_k I)x \\ &= (A - \mu_1 I)\varphi = 0, \end{aligned} \tag{1.3.8}$$

其中

$$\varphi = (A - \mu_2 I)\cdots(A - \mu_k I)x \neq 0. \tag{1.3.9}$$

注意, $\varphi \neq 0$ 是由 k 的最小性推出的. (1.3.8) 表明, $A\varphi = \mu_1\varphi$, 即 φ 是 A 的对应于特征值 μ_1 的特征向量. 从 (1.3.9) 得

$$\varphi = A^{k-1}x + \left(-\sum_{j=2}^{k}\mu_j\right)A^{k-2}x + \cdots + (-1)^{k-1}\prod_{j=2}^{k}\mu_j x.$$

这就证明了

$$\varphi \in \mathscr{M}(A^{k-1}x, A^{k-2}x, \cdots, x).$$

引理得证.

定理 1.3.5 设 A, B 为 n 阶实对称阵, 则存在正交阵 Q, 使得 $Q'AQ$ 与 $Q'BQ$ 为对角阵, 当且仅当 $AB = BA$.

证明 必要性的证明是容易的, 下面证明充分性. 设 φ_1 为 B 的对应于特征值 λ 的特征向量, 于是, $B\varphi_1 = \lambda\varphi_1$, 用 A^k 左乘两边, 利用 $AB = BA$ 得

$$A^kB\varphi_1 = \lambda A^k\varphi_1 = BA^k\varphi_1, \quad k = 1, 2, \cdots,$$

于是

$$B(A^k\varphi_1) = \lambda(A^k\varphi_1), \quad k = 1, 2, \cdots.$$

这表明: $\varphi_1, A\varphi_1, A^2\varphi_1, \cdots$ 都是 B 的对应于同一个特征值 λ 的特征向量. 依引理 1.3.1, 存在 A 的一个特征向量 q_1, 使得

$$q_1 \in \mathscr{M}(\varphi_1, A\varphi_1, A^2\varphi_1, \cdots).$$

因为 $\mathscr{M}(\varphi_1, A\varphi_1, A^2\varphi_1, \cdots)$ 是 B 的一个特征子空间, 于是 q_1 是 A, B 的一个公共特征向量.

选择 B 的另一个特征向量 φ_2, 且满足 $\varphi_2 \perp q_1$. 用 φ_2 代替 φ_1, 重复上面的讨论, 我们可以找到 A 与 B 的另一个公共特征向量 q_2, 且 $q_2 \perp q_1$. 继续这个过程, 我们可以找到 A 与 B 的 n 个公共特征向量 $q_1, \cdots, q_n$ 满足 $q_j'q_k = 0$. 不妨假设它们的长度都等于 1, 于是 $Q = (q_1, \cdots, q_n)$ 就是所求的正交阵. 定理得证.

注 4 从本定理知, 可交换的两个 n 阶实对称阵具有 n 个两两正交的公共特征向量, 但它们的特征值不必相同.

定理 1.3.6 设 A, B 为两个 n 阶实半正定阵, 则存在可逆阵 Q, 使得 $Q'AQ$ 和 $Q'BQ$ 皆为对角阵.

证明 设 $t = r(A + B)$. 则存在可逆阵 P, 使得

$$P'(A + B)P = \begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}. \tag{1.3.10}$$

将 $P'BP$ 分块为

$$P'BP = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \tag{1.3.11}$$

其中 M_{11} 为 $t \times t$ 矩阵, 因为 $P'(A + B)P \geqslant P'BP$(这里 $C \geqslant D$ 表示 $C - D \geqslant 0$), 于是从 (1.3.10) 和 (1.3.11) 可以推出

$$M_{ij} = 0, \qquad \text{当 } i + j \geqslant 3 \text{ 时},$$

即

$$P'BP = \begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix}. \tag{1.3.12}$$

注意到 M_{11} 仍为对称阵, 应用定理 1.3.1(2) 知, 存在正交阵 Q_1, 使得

$$Q_1^{-}M_{11}Q_1 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_t \end{pmatrix}.$$

记

$$Q = P\begin{pmatrix} Q_1 & 0 \\ 0 & I_{n-t} \end{pmatrix},$$

则 Q 为可逆阵, 且有

$$Q'BQ = \mathrm{diag}(\lambda_1, \cdots, \lambda_t, 0, \cdots, 0),$$
$$Q'AQ = \mathrm{diag}(1-\lambda_1, \cdots, 1-\lambda_t, 0, \cdots, 0).$$

定理证毕.

§1.4　Hermite 阵

设 $A = a_{ij}$ 为 n 阶方阵, 记 $A^* = \bar{A}'$, 即取共轭同时又转置. 若 $A^* = A$, 则称 A 是一个 Hermite 阵. 当 A 为实矩阵时, Hermite 阵就是实对称阵.

Hermite 阵具有许多类似于实对称阵的重要性质.

定理 1.4.1　设 A 为 n 阶 Hermite 阵, 则

(1) A 的所有特征值都是实数;

(2) 存在一个酉阵 U, 即 U 满足 $U^*U = I_n$, 使得

$$U^*AU = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \tag{1.4.1}$$

即 Hermite 阵一定酉相似于对角阵.

注 1　从 (1.4.1) 知

$$A \sim \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

所以, $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值, 若记 $U = (u_1, \cdots, u_n)$, 则从 (1.4.1) 可推出

$$AU = U\mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

即 $Au_j = \lambda_j u_j, j = 1, \cdots, n$. 所以, $u_1, \cdots, u_n$ 为 A 的 n 个标准正交化特征向量.

注 2　记 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 则 (1.4.1) 可改写为

$$A = U\Lambda U^* = \sum_{j=1}^{n} \lambda_j u_j u_j^*,$$

称为 A 的谱分解.

由 (1.4.1) 容易得到如下推论.

推论 1.4.1 设 A 为 n 阶 Hermite 阵, x 为 $n\times 1$ 向量. 那么存在酉阵 U, 使得酉变换 $y=U^*x$ 把二次型 x^*Ax 化为平方和

$$x^*Ax=y^*\Lambda y=\lambda_1|y_1|^2+\cdots+\lambda_n|y_n|^2,$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$.

因此, 完全类似于上节, 我们可以定义正定和半正定的 Hermite 阵.

设 A 为 n 阶 Hermite 阵, 若对任意 $x\in C^n, x^*Ax\geqslant 0$, 则称 A 为半正定的, 记为 $A\geqslant 0$. 若进一步, $x^*Ax=0$ 当且仅当 $x=0$, 则称 A 为正定的, 记为 $A>0$.

类似于定理 1.3.2 和 1.3.3, 我们有

定理 1.4.2 设 A 为 n 阶 Hermite 阵, 则 $A\geqslant 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值为非负;

(2) 存在一个 Hermite 阵 B, 使得 $A=B^2$;

(3) 存在 $t\times n$ 矩阵 B, 其中 $t=r(A)$, 使得 $A=B^*B$;

(4) 对任一复方阵 P, $P^*AP\geqslant 0$;

(5) A 的所有主子式为非负.

定理 1.4.3 设 A 为 n 阶 Hermite 阵, 则 $A>0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值为正数;

(2) 存在一个可逆 Hermite 阵 B, 使得 $A=B^2$;

(3) 存在可逆复方阵 B, 使得 $A=B^*B$;

(4) 对任一可逆复方阵 P, $P^*AP>0$;

(5) A 的所有主子式为正数;

(6) A 的所有顺序主子式为正数.

注 3 对半正定的 Hermite 阵 A, 也存在着半正定的平方根阵 $A^{1/2}$, 使得 $(A^{1/2})^2=A$. 事实上, 根据定理 1.4.1, 存在酉阵 U, 使得 $A=U\Lambda U^*$, 这里 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n), \lambda_i\geqslant 0, i=1,\cdots,n$ 为 A 的特征值, 记 $\Lambda^{1/2}=\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})$, 则 A 可以改写为

$$A=U\Lambda^{1/2}U^*U\Lambda^{1/2}U^*=B^2,$$

其中 $B=U\Lambda^{1/2}U^*\geqslant 0$, 就是 A 的平方根阵 $A^{1/2}$. 显然当 $A>0$ 时, $B=A^{1/2}>0$.

类似于定理 1.3.4~1.3.6, 我们有如下几个定理. 上节的证明过程照搬过来, 将其中的 "正交阵"、"实对称阵" 和 "Q'" 分别改为 "酉阵"、"Hermite 阵" 和 "Q^*" .

定理 1.4.4 设 A, B 为两个 n 阶 Hermite 阵, 且 $B>0$. 则存在可逆阵 Q, 使得

$$A=Q^*\Lambda Q,$$
$$B=Q^*Q,$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),\lambda_j(j=1,\cdots,n)$ 为 AB^{-1} 的特征值.

定理 1.4.5　设 A, B 为两个 n 阶 Hermite 阵. 则存在酉阵 U, 使得 U^*AU 和 U^*BU 为对角阵, 当且仅当 $AB=BA$.

定理 1.4.6　设 A, B 为两个同阶半正定 Hermite 阵, 则存在可逆阵 Q, 使得 Q^*AQ 与 Q^*BQ 皆为对角阵.

§1.5 矩 阵 分 解

所谓矩阵分解, 就是将一个矩阵写成从某种意义上讲比较简单或对它的性质比较熟悉的若干矩阵的乘积. 例如, 在 §1.3 中, 实对称阵 A 被分解为 $A=\Phi\Lambda\Phi'$, 这里 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 是对角阵, 有比较简单的形式, 而 Φ 是一个正交阵, 具有许多重要性质. 其中最重要的一条性质是不改变向量长度, 也就是 $\|\Phi x\|=\|x\|$. 实对称阵的这种分解, 对研究实对称阵的性质带来很大方便.

本节我们叙述一般矩阵的几种重要分解.

定理 1.5.1 (Schmidt 三角化分解)　设 A 为 $m\times n$ 复矩阵, $r(A)=n$. 则存在 $n\times n$ 上三角阵 R 和 $m\times n$ 的矩阵 Q, $Q^*Q=I_n$, 使得 $A=QR$.

证明　设 $A=(a_1,\cdots,a_n)$, 因为 $a_1,\cdots,a_n$ 线性无关, 对它们施用 Schmidt 正交化程序, 即得所要结论.

定理 1.5.2 (秩分解)　设 A 为 $m\times n$ 复方阵, 则存在两个可逆阵 $P_{m\times m}$ 和 $Q_{n\times n}$, 使得

$$A=P\begin{pmatrix} I_t & 0\\ 0 & 0\end{pmatrix}Q, \tag{1.5.1}$$

这里 $t=r(A)$.

证明要点如下. 矩阵 A 经过行向量的如下三种初等变换:

(1) 交换两行的位置;

(2) 用一个非零数去乘某一行;

(3) 将某一行乘一个数加到另一行,

可化为行约简梯形, 成为满足下列四条的矩阵:

(1) 非零行的第一个非零元素为 1, 以下称为主导元素 1;

(2) 各列中, 若含有主导元素 1, 则其余元素全为零;

(3) 所有元素为零的行排在矩阵的下部;

(4) 主导元素 1 排成阶梯形, 即若 t 个行是非零行, 且第 j 行的主导元素 1 位于第 k_i 列, $j=1,\cdots,t$, 则 $k_1<k_2<\cdots<k_t$.

因为对矩阵 A 施以一系列三种行初等变换相当于用一个可逆阵左乘矩阵A, 于是上面我们证明了, 存在可逆阵 B, 使得 BA 成行约简梯形. 对行约简梯形 BA 再

施以列的三种初等变换, 这又相当于用一个可逆阵 C 右乘 BA, 此时 BA 化为

$$BAC = \begin{pmatrix} I_t & 0, \\ 0 & 0 \end{pmatrix},$$

因为 B, C 皆可逆, 取 $P = B^{-1}, Q = C^{-1}$, 得到 (1.5.1).

注 1 秩分解 (1.5.1) 也称为 A 的等价标准形.

定理 1.5.3 (满秩分解) 设 A 为 $m \times n$ 复阵, $r(A) = t$. 则存在 $m \times t$ 和 $t \times n$ 且秩为 t 的矩阵 B 和 C, 使得

$$A = BC. \tag{1.5.2}$$

证明 应用 A 的秩分解 (1.5.1), 将 P 和 Q 分块为

$$P = (P_1 \vdots P_2), \quad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix},$$

这里 P_1 和 Q_1 分别为 $m \times t$ 和 $t \times n$ 矩阵, 则 $A = P_1 Q_1$. 明所欲证.

定理 1.5.4 (奇异值分解) 设 A 为 $m \times n$ 秩为 t 的复矩阵, 则存在两个酉阵 $U_{m\times m}$ 和 $V_{n\times n}$, 使得

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*, \tag{1.5.3}$$

其中 $\Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_t), \sigma_i > 0, i = 1, \cdots, t$. $\lambda_1 = \sigma_1^2, \cdots, \lambda_t = \sigma_t^2$ 为 A^*A 的非零特征值.

证明 设 $u_1, \cdots, u_t$ 为 A^*A 的对应于非零特征值 $\sigma_1^2, \cdots, \sigma_t^2$ 的标准正交化特征向量. 记 $v_j = \dfrac{1}{\sigma_j} A^* u_j, j = 1, \cdots, t$. 容易验证, $v_1, \cdots, v_t$ 是标准正交化的.

将 $u_1, \cdots, u_t$ 和 $v_1, \cdots, v_t$ 分别扩充为 C^m 和 C^n 的标准正交基

$$u_1, \cdots, u_t, u_{t+1}, \cdots, u_m; \tag{1.5.4}$$

$$v_1, \cdots, v_t, v_{t+1}, \cdots, v_n, \tag{1.5.5}$$

并记以向量组 (1.5.4) 和 (1.5.5) 分别为列向量的矩阵为 U 和 V, 则 $U^*U = I_m, V^*V = I_n$, 且

$$\begin{aligned} A &= UU^*A \\ &= (u_1 u_1^* + \cdots + u_t u_t^*)A \\ &= \sigma_1 u_1 v_1^* + \cdots + \sigma_t u_t v_t^* \\ &= U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*. \end{aligned}$$

证毕.

注 2 $\sigma_1=\lambda_1^{1/2},\cdots,\sigma_t=\lambda_t^{1/2}$ 称为 A 的奇异值. 有时, 为方便计, 也常把 A^*A 的所有特征值的平方根 $\lambda_1^{1/2},\cdots,\lambda_n^{1/2}$ 称为 A 的奇异值. 若 $r(A)=t$, 则后面的 $n-t$ 个奇异值 $\lambda_{++1}^{1/2}=\cdots=\lambda_n^{1/2}=0$.

注 3 由奇异值分解 (1.5.3) 可推得

$$AA^*=U\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}U^*,$$

$$A^*A=V\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}V^*.$$

这表明 U 的 m 个列向量是 AA^* 的标准正交化特征向量, 而 V 的 n 个列向量是 A^*A 的标准正交化特征向量.

特别地, 当 $AA^*=A^*A$ 时, 我们可以选取 $U=V$.

设 A 为复方阵, 若 $AA^*=A^*A$, 则称 A 为复规范阵. 根据上面的讨论, 对复规范阵 A, 奇异值分解 (1.5.3) 中, 可取 $U=V$. 这就证明了如下定理.

定理 1.5.5 设 A 为 n 阶复规范阵, 则存在酉阵 U, 使得

$$A=U^*\Lambda U,$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),\lambda_1,\cdots,\lambda_n$ 为 A 的特征值.

定理 1.5.6 (Jordan 分解) 设 A 为 n 阶方阵, 则存在复的可逆方阵 P, 使得

$$A=P\mathrm{diag}(\Lambda_1,\Lambda_2,\cdots,\Lambda_k)P^{-1},$$

这里

$$\Lambda_j=\begin{pmatrix}\lambda_j & 1 & & 0 & \cdots & 0\\ & & & & \ddots & \vdots\\ & \lambda_j & & 1 & & 0\\ & & \ddots & & \ddots & \\ & & & \ddots & & 1\\ & & & & \ddots & \lambda_j\end{pmatrix},\quad j=1,\cdots,k$$

为对应于 A 的特征值 λ_j 的 Jordan 块.

定理 1.5.7 (Schur 分解) 设 A 为 $n\times n$ 复方阵, 则存在酉阵 U, 使得 $A=ULU^*$, 其中 L 为上 (下) 三角阵, L 的对角元为 A 的特征值. 若 A 是 $n\times n$ 实方阵, 且它的特征值也为实数, 则 U 为正交阵.

证明 设 A 的特征值为 $\lambda_1,\cdots,\lambda_n, x_1\in C^n, x_1$ 为对应于 λ_1 的单位特征向量. 将 x_1 扩充为 C^n 的一组基：$x_1,y_2,\cdots,y_n$, 应用 Schmidt 正交化程序，将这些向量变换为 C^n 的一组标准正交基 $x_1,z_2,\cdots,z_n$. 定义 n 阶酉阵 $U_1=(x_1,z_2,\cdots,z_n)$, 则有

$$U_1^*AU_1=\begin{pmatrix}\lambda_1 & * \\ 0 & A_1\end{pmatrix},$$

其中 A_1 为 $n-1$ 阶方阵, 它的特征值为 $\lambda_2,\cdots,\lambda_n$. 设 $x_2\in C^{n-1}$, 且是 A_1 的对应于特征值 λ_2 的单位特征向量. 重复上面的方法可以得到 $n-1$ 阶酉阵 U_2, 使得

$$U_2^*A_1U_2=\begin{pmatrix}\lambda_2 & * \\ 0 & A_2\end{pmatrix}.$$

定义

$$\widetilde{U}_2=\begin{pmatrix}1 & 0 \\ 0 & U_2\end{pmatrix},$$

则 $\widetilde{U}_2$ 和 $U_1\widetilde{U}_2$ 都是酉阵, 且使

$$\widetilde{U}_2^*U_1^*AU_1\widetilde{U}_2=\left(\begin{array}{cc:c}\lambda_1 & * & * \\ 0 & \lambda_2 & \\ \hdashline 0 & & A_2\end{array}\right),$$

这里 A_2 为 $n-2$ 阶方阵, 它的特征值为 $\lambda_3,\cdots,\lambda_n$. 重复上述过程, 我们会构造出 n 阶酉阵 $U_1,\widetilde{U}_2,\cdots,\widetilde{U}_{n-1}$. 记 $U=U_1\widetilde{U}_2\cdots\widetilde{U}_{n-1}$, 使有 U^*AU 为上三角阵.

如果我们首先对 A^* 作分解：$A^*=ULU^*$, 这里 U 为酉阵, L 为上三角阵, 则 $A=UL^*U^*$, 此时 L^* 就是下三角阵了.

最后, 若 A 的元素和所有特征值都是实数, 那么它的所有特征向量都可取为实向量, 此时酉阵 U 便为正交阵. 定理证毕.

§1.6 矩阵的范数

为了度量一个向量的“大小”, 在 §1.1 我们引进了向量的长度. 对于一个 $m\times n$ 矩阵, 我们自然也可以把它看成 mn 维向量, 按向量的办法定义它的“长度”. 但是, 下面我们用更一般的途径引进度量矩阵“大小”的量, 这就是矩阵的范数.

定义 1.6.1 设 A 为 $m\times n$ 复矩阵, 定义非负实函数 $\|A\|$. 若它满足下列性质：

(1) 正定性：$\|A\|\geqslant 0$, 且 $\|A\|=0 \Longleftrightarrow A=0$;

(2) 齐次性：$\|\alpha A\|=|\alpha|\|A\|, \alpha$ 为一数;

(3) 三角不等式：$\|A+B\| \leqslant \|A\|+\|B\|$,
则称 $\|A\|$ 为 A 的范数. 若对任意的 $A_{m\times n}$ 和 $B_{n\times l}$, 有

$$\|AB\| \leqslant \|A\|\|B\|, \tag{1.6.1}$$

则称该范数是相容的.

例 1.6.1 (欧氏范数)　设 $A=(a_{ij})_{m\times n}$. 定义

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}. \tag{1.6.2}$$

则它是一种范数, 称为矩阵 A 的 Euclid 范数, 简称欧氏范数, 或称 Frobenius 范数. $\|A\|_F$ 也可改写为如下形式

$$\|A\|_F = (\mathrm{tr}A^*A)^{1/2}, \tag{1.6.3}$$

从此式可以看出, 这是把 A 按列先后排列成 $mn\times 1$ 向量后, 取其欧氏长度得到的.

因为

$$\begin{aligned}\|AB\|_F^2 &= \mathrm{tr}(B^*A^*AB)\\ &\leqslant \lambda_1(A^*A)\mathrm{tr}(B^*B) \leqslant (\mathrm{tr}A^*A)(\mathrm{tr}B^*B)\\ &= \|A\|_F^2\|B\|_F^2,\end{aligned}$$

其中 $\lambda_1(D)$ 表示 D 的最大特征值, 所以, 欧氏范数是相容的.

例 1.6.2 (谱范数)

$$\|A\|_2 = \lambda_1^{1/2}(A^*A),$$

即矩阵 A 的谱范数是 A^*A 的最大特征值的平方根. 谱范数也是相容范数. 事实上

$$\begin{aligned}\|AB\|_2^2 &= \lambda_1(B^*A^*AB)\\ &\leqslant \lambda_1(B^*\cdot\lambda_1(A^*A)I\cdot B)\\ &\leqslant \lambda_1(A^*A)\cdot\lambda_1(B^*B)\\ &= \|A\|_2^2\|B\|_2^2,\end{aligned}$$

这里, 在两个不等号处应用了特征值的单调性：若 $A\geqslant 0, B\geqslant 0, A\geqslant B$, 则 $\lambda_1(A) \geqslant \lambda_1(B)$(证明见推论 4.4.1).

根据上面的性质 (3) 还可以证明如下不等式

$$|\|A\|-\|B\|| \leqslant \|A-B\|. \tag{1.6.4}$$

事实上,

$$\|A\| = \|A-B+B\| \leqslant \|A-B\|+\|B\|,$$

所以

$$\|A\| - \|B\| \leqslant \|A - B\|. \tag{1.6.5}$$

同理可证

$$\|B\| - \|A\| \leqslant \|B - A\| = \|A - B\|. \tag{1.6.6}$$

由 (1.6.5) 和 (1.6.6) 便有 (1.6.4).

下面我们证明有关谱范数的一个重要定理.

定理 1.6.1 设 $\|A\|_2 < 1$, 则 $I - A$ 可逆, 且

$$\|(I - A)^{-1}\|_2 \leqslant \frac{1}{1 - \|A\|_2}, \tag{1.6.7}$$

$$\|I - (I - A)^{-1}\|_2 \leqslant \frac{\|A\|_2}{1 - \|A\|_2}. \tag{1.6.8}$$

证明 对任意 $n \times 1$ 向量 x, $\|x\|_2$ 就等于它的欧氏长度 $\|x\|$, 应用相容性及 (1.6.4), 得

$$\begin{aligned} \|(I - A)x_2\| &= \|x - Ax\|_2 \geqslant \|x\|_2 - \|Ax\|_2 \\ &\geqslant \|x\|_2 - \|A\|_2\|x\|_2 \\ &= (1 - \|A\|_2)\|x\|_2. \end{aligned}$$

可见, 当 $x \neq 0$ 时, $(I - A)x \neq 0$, 即 $(I - A)x = 0$ 只有零解, 故 $I - A$ 可逆.

因为

$$(I - A)^{-1} = I + A(I - A)^{-1}, \tag{1.6.9}$$

应用谱范数的相容性及三角不等式, 得

$$\begin{aligned} \|(I - A)^{-1}\|_2 &\leqslant \|I\|_2 + \|A(I - A)^{-1}\| \\ &\leqslant 1 + \|A\|_2\|(I - A)^{-1}\|_2, \end{aligned}$$

由此立得 (1.6.7). 另一方面, 将 (1.6.9) 变形为

$$I - (I - A)^{-1} = -A(I - A)^{-1}.$$

应用 (1.6.7) 得

$$\|I - (I - A)^{-1}\|_2 \leqslant \|A\|_2\|(I - A)^{-1}\|_2 \leqslant \frac{\|A\|_2}{1 - \|A\|_2}.$$

证毕.

对于一个矩阵 $A_{m\times n}$, 它的范数有许多种. 因此根据范数所具有的性质, 我们可以对它们进行分类. 例如, 若 $\|A\|$ 满足如下附加性质

$$\|U_1 A U_2\| = \|A\|,$$

这里 U_1 和 U_2 分别为 m 阶和 n 阶酉阵, 则 $\|A\|$ 称为酉不变范数. 容易验证, 前面两个例子中引进的欧氏范数和谱范数都是酉不变范数. 但是下面两种范数就不是酉不变的.

$$\|A\|_1 = \max_{1\leqslant i\leqslant n}\sum_{i=1}^{m}|a_{ij}|,$$

$$\|A\|_\infty = \max_{1\leqslant j\leqslant n}\sum_{j=1}^{n}|a_{ij}|,$$

它们分别叫做列和范数和行和范数. 关于这些范数的进一步讨论, 读者可参阅相关文献 (孙继广 1987).

§1.7 广义逆矩阵

设 A 为 $n\times n$ 方阵且满秩, 众所周知, 存在 A 的唯一逆矩阵, 记为 A^{-1}, 满足

$$AA^{-1} = A^{-1}A = I_n.$$

此时, 线性方程组 $Ax = b$ 有唯一解 $x = A^{-1}b$.

如果 A 不是满秩的或它根本就不是方阵, 这时相容线性方程组 $Ax = b$ 的解应该如何表示呢? 这就需要将经典的逆矩阵概念加以推广. 为此, 人们提出了 “广义逆矩阵” 的概念. 近四十年来, 广义逆矩阵的理论和计算方法有了迅速的发展, 已成为矩阵理论研究和应用的一个不可缺少的工具.

记 A 为 $m\times n$ 矩阵, 任意 $n\times m$ 矩阵 X, 若满足

$$AXA = A; \tag{1.7.1}$$

$$XAX = X; \tag{1.7.2}$$

$$(AX)^* = AX; \tag{1.7.3}$$

$$(XA)^* = XA \tag{1.7.4}$$

中的一个或几个条件, 都称为 A 的广义逆矩阵. 上面的四个方程称为 Moore-Penrose 方程. 若 X 满足这四个方程中的第 $i, j, \cdots, l$ 个, 则称 X 为 A 的 $\{i, j, \cdots, l\}$ 广义逆, 记为 $A^{(i,j,\cdots,l)}$. 但比较重要的是下面五种广义逆:

(1) $A^{(1)}$, 就是只满足方程 (1.7.1) 的广义逆, 以后简记为A^-.

(2) $A^{(1,2,3,4)}$, 就是满足 (1.7.1) ~ (1.7.4) 全部四个方程的广义逆, 也被称为 Moore-Penrose 广义逆, 以后简记为 A^+;

(3) $A^{(1,2)}$, 也就是只满足前两个条件的广义逆, 称为 A 的自反广义逆;

(4) $A^{(1,3)}$, 称为最小二乘广义逆, 这个名称的渊源是: 对任意的相容或不相容(即矛盾) 线性方程组 $Ax=b, x=A^{(1,3)}b$ 是该方程组的最小二乘解;

(5) $A^{(1,4)}$, 称为最小范数广义逆, 这是因为 $x=A^{(1,4)}b$ 是相容线性方程组 $Ax=b$ 解集中范数最小者.

本书后面的讨论只涉及广义逆 A^- 和 A^+, 为简单计, 我们以后用术语 "广义逆" 专指 A^-, 而 A^+ 总是叫做 Moore-Penrose 广义逆. 现在进一步讨论这两种广义逆的性质.

定理 1.7.1 (1) 设 A 为 $m\times n$ 阵, P 和 Q 分别为 $m\times m$ 和 $n\times n$ 可逆阵, 则

$$(PAQ)^-=Q^{-1}A^-P^{-1}; \tag{1.7.5}$$

(2) 记 I_t 为 t 阶单位阵, 则

$$\begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}^-=\begin{pmatrix} I_t & B \\ C & D \end{pmatrix},$$

其中 B, C 和 D 为适当阶数的任意矩阵;

(3) 设 A 为 $m\times n$ 阵, $r(A)=t$, 且有秩分解

$$A=P\begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}Q,$$

这里 P 和 Q 分别为 $m\times m, n\times n$ 可逆阵, 则

$$A^-=Q^{-1}\begin{pmatrix} I_t & B \\ C & D \end{pmatrix}P^{-1}.$$

其中 B, C 和 D 为适当阶数的任意矩阵.

(1) 和 (2) 的证明是容易的, (3) 是 (1) 和 (2) 的推论.

下面是广义逆 A^- 的表示定理, 即通过一个特定的广义逆把 A 的所有广义逆表征出来.

定理 1.7.2 设 A 为 $m\times n$ 阵, A^- 为其任一特定广义逆. 则 A 的任一广义逆 X 可表为如下形式

(1) $X=A^-+U-A^-AUAA^-$; (1.7.6)

(2) $X=A^-+V(I_m-AA^-)+(I_n-A^-A)U$, (1.7.7)

这里 U 和 V 皆为 $n\times m$ 的任意阵.

证明 容易验证, (1.7.6) 和 (1.7.7) 给出的 X 都满足 $AXA=A$. 另一方面, 对 $AXA=A$ 的任一解 X, 若取 $U=X-A^-$, 则它具有形式 (1.7.6); 若取 $V=X-A^-, U=XAA^-$ 就具有形式 (1.7.7). 证毕.

由定义和定理 1.7.1 容易推出 A^- 的下列性质.

定理 1.7.3 (1) $(A^-)^* = (A^*)^-$;

(2) 若 A 可逆, 则 A^- 唯一, 且 $A^- = A^{-1}$;

(3) 若记

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0, \\ 0, & \lambda = 0, \end{cases}$$

则 $(\lambda A)^- = \lambda^+ A^-$;

(4) $r(A^-) \geqslant r(A)$;

(5) AA^- 和 A^-A 都是幂等阵, 且 $r(AA^-) = r(A^-A) = r(A)$. 这里一个方阵 B 被称为幂等阵当且仅当 $B^2 = B$. (关于幂等阵的详细讨论见 §1.8.)

一般说来, 矩阵 BA^-C 与其中广义逆 A^- 的选择有关, 对不同的广义逆 A^-, BA^-C 就不同. 但是若 B, C 与 A 有一定关系时, BA^-C 就与 A^- 的选择无关. 这个性质很重要, 广义逆在数理统计中的很多应用都与此有关.

定理 1.7.4 设 $B \neq 0, C \neq 0$, 则 BA^-C 与 A^- 的选择无关, 当且仅当 $\mathscr{M}(C) \subset \mathscr{M}(A), \mathscr{M}(B') \subset \mathscr{M}(A')$.

证明 先证充分性. 由假设知, 存在矩阵 X 和 Y, 使得 $C = AX, B' = A'Y$. 于是

$$BA^-C = Y'AA^-AX = Y'AX,$$

右端与 A^- 无关.

为证必要性, 应用广义逆的表示 (1.7.6), 得

$$BA^-C = BA^-C + BUC - BA^-AUAA^-C, \text{ 对一切 } U, \tag{1.7.8}$$

于是

$$BUC - BA^-AUAA^-C = 0, \text{ 对一切 } U.$$

特别取 $U = A^-AZ$, 其中 Z 为任意阵, 则有

$$BA^-AZ(C - AA^-C) = 0.$$

从 Z 的任意性可推知, 或

$$BA^-A = 0, \tag{1.7.9}$$

或

$$C = AA^-C. \tag{1.7.10}$$

若 (1.7.9) 成立, 那么从 (1.7.8) 得

$$BUC = 0, \text{ 对一切 } U.$$

这表明 $B=0$ 或 $C=0$, 这与原假设矛盾. 于是, (1.7.10) 成立, 也就是 $\mathscr{M}(C)\subset\mathscr{M}(A)$. 类似地可证 $\mathscr{M}(B')\subset\mathscr{M}(A')$. 必要性得证. 定理证毕.

推论 1.7.1 设 $B\neq 0$, 则 $B'A^-B$ 与 A 的选择无关当且仅当 $\mathscr{M}(B)\subset\mathscr{M}(A)$.

定理 1.7.5 对任意矩阵 A, 有

(1) $A(A^*A)^-A^*$ 与广义逆 $(A^*A)^-$ 的选择无关;

(2) $A(A^*A)^-A^*A=A, A^*A(A^*A)^-A^*=A^*$.

证明 (1) 因为 $\mathscr{M}(A^*)=\mathscr{M}(A^*A)$, 故存在矩阵 X, 使得 $A^*=A^*AX$, 于是

$$\begin{aligned}A(A^*A)^-A^* &= X^*A^*A(A^*A)^-A^*AX\\ &= X^*A^*AX,\end{aligned}$$

右端与 $(A^*A)^-$ 选择无关.

(2) 记 $D=A(A^*A)^-A^*A-A$, 可直接验证 $D^*D=0$, 于是 $D=0$. 类似地, 可证第二式.

前面我们讨论的是广义逆 A^- 的一些重要性质. Moore-Penrose 广义逆 A^+ 作为一个特殊的 A^-, 除了具有 A^- 的全部性质外, 还具有一些特殊性质.

我们先给出 A^+ 的构造性定理.

定理 1.7.6 设 $m\times n$ 矩阵 A 的奇异值分解为

$$A=P\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}Q^*,$$

其中, P 和 Q 分别为 $m\times m$ 和 $n\times n$ 酉阵, $\Sigma=\operatorname{diag}(\sigma_1,\cdots,\sigma_t), t=r(A)$. 则

$$A^+=Q\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{pmatrix}P^*.$$

证明是容易的, 留作练习.

与其他广义逆不同, A^+ 总是唯一的. 此即

定理 1.7.7 对任意矩阵 A, A^+ 总是唯一的.

证明 在定理 1.7.6, 我们已构造出了 A^+. 设 X 满足 Moore-Penrose 方程, 我们要证明 $X=A^+$. 事实上, 反复利用 (1.7.1)~(1.7.4), 我们有

$$\begin{aligned}X &= XAX=X(AX)^*=XX^*A^*=XX^*(AA^+A)^*\\ &= X(AX)^*(AA^+)^*=XAXAA^+=XAA^+=(XA)^*A^+AA^+\\ &= A^*X^*A^+AA^+=A^*X^*A^*(A^+)^*A^+=(AXA)^*(A^+)^*A^+\\ &= A^*(A^+)^*A^+=(A^+A)^*A^+=A^+AA^+=A^+.\end{aligned}$$

证毕.

下面的定理从满秩分解给出 A^+ 的另一种构造.

定理 1.7.8　设 $m\times n$ 矩阵 A 的秩为 $t(t>0)$, 且满秩分解为

$$A=FG, \tag{1.7.11}$$

这里 F 和 G 分别为 $m\times t$ 和 $t\times n$ 矩阵, 且 $r(F)=r(G)=r(A)=t$. 则

$$A^+=G^*(F^*AG^*)^{-1}F^*. \tag{1.7.12}$$

证明　我们首先证明 F^*AG^* 为可逆阵. 从 (1.7.11) 我们有

$$F^*AG^*=F^*FGG^*.$$

注意到 $r(F^*F)=r(F)=t, r(GG^*)=r(G)=t$, 且 F^*F 和 GG^* 都是 t 阶方阵, 所以两者都可逆, 故 F^*AG^* 可逆, 于是

$$(F^*AG^*)^{-1}=(GG^*)^{-1}(F^*F)^{-1},$$

$$G^*(F^*AG^*)^{-1}F^*=G^*(GG^*)^{-1}(F^*F)^{-1}F^*.$$

容易验证, 上式右端满足 Moore-Penrose 方程. 定理证毕.

下面的推论是定理 1.7.8 的两个特例.

推论 1.7.2　(1) 假设 $a\neq 0, b\neq 0$, 则

$$(ab^*)^+=\frac{ba^*}{a^*a\cdot b^*b}=\frac{ba^*}{\|a\|^2\cdot\|b\|^2};$$

(2) 若 $a\neq 0$, 则

$$a^+=\frac{a^*}{a^*a}=\frac{a^*}{\|a\|^2}.$$

证明　在定理中, 若 $t=1$, 便得到 (1). (2) 是 (1) 的 $b=1$ 的特例.

一般说来, 对任意可逆阵 P 和 Q, 关系式 $(PAQ)^+=Q^{-1}A^+P^{-1}$ 不必成立, 但是我们有下面的结论.

定理 1.7.9　设 A 为 $m\times n$ 矩阵, 有

(1) 若 P 和 Q 分别为 $m\times m$ 和 $n\times n$ 酉阵, 则

$$(PAQ)^+=Q^{-1}A^+P^{-1}.$$

(2) 若 P 和 Q 分别为 $k\times m$ 和 $l\times m$ 矩阵, 满足 $P^*P=I_m, Q^*Q=I_n$, 即 P 和 Q 的列向量是标准正交的向量组. 则

$$(PAQ^*)^+=QA^+P^*.$$

定理的证明是简单的.

定理 1.7.10 对任意矩阵 A, 有

(1) $(A^+)^+ = A$;

(2) $(A^+)^* = (A^*)^+, (A^+)' = (A')^+$;

(3) $A^+ = (A^*A)^+A^* = A^*(AA^*)^+$;

(4) $I \geqslant A^+A$, 这里记号 $B \geqslant C$ 表示 $B - C \geqslant 0$, 即 $B - C$ 为半正定阵.

下面我们讨论分块矩阵的广义逆. 因为证明分块矩阵的普通逆矩阵的思路和处理技巧可以直接应用到广义逆的情况, 所以我们的讨论先从可逆阵开始.

设

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{1.7.13}$$

并记

$$A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}, \tag{1.7.14}$$

$$A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}. \tag{1.7.15}$$

定理 1.7.11 假设 A 可逆且分块为 (1.7.13), 有

(1) 若 $\det A_{11} \neq 0$, 则 $\det A_{22.1} \neq 0$, 且

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix}; \tag{1.7.16}$$

(2) 若 $\det A_{22} \neq 0$, 则 $\det A_{11.2} \neq 0$, 且

$$A^{-} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}. \tag{1.7.17}$$

证明 (1.7.17) 和 (1.7.16) 的证明相类似, 我们只证 (1.7.16).

因为 $\det A_{11} \neq 0$, 则

$$\begin{aligned} &\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix} \end{aligned} \tag{1.7.18}$$

由此可推得 $\det A_{22.1} \neq 0$. 将上式两边求逆矩阵, 整理后得

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}.$$

右边三个矩阵相乘就得到 (1.7.16). 证毕.

当 A^{-1} 不存在时, 即便 A_{11} 或 A_{22} 可逆, $A_{11.2}$ 和 $A_{22.1}$ 也不必可逆, 此时关于 A^- 的表达式 (1.7.16) 和 (1.7.17) 仍然成立, 但需要将所有 $A_{22.1}^{-1}$ 和 $A_{11.2}^{-1}$ 换为 $A_{22.1}^-$ 和 $A_{11.2}^-$. 此即

定理 1.7.12　假设 A 分块为 (1.7.13), 有

(1) 若 A_{11}^{-1} 存在, 则

$$A^- = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^-A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^- \\ -A_{22.1}^-A_{21}A_{11}^{-1} & A_{22.1}^- \end{pmatrix}; \tag{1.7.19}$$

(2) 若 A_{22}^{-1} 存在, 则

$$A^- = \begin{pmatrix} A_{11.2}^- & -A_{11.2}^-A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^- & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^-A_{12}A_{22}^{-1} \end{pmatrix}. \tag{1.7.20}$$

证明　我们只证明 (1.7.19), (1.7.20) 的证明类似.

首先注意到, 当 A_{11}^{-1} 存在时, (1.7.18) 仍然成立. 利用 (1.7.5), 我们得到

$$A^- = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}^-\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}. \tag{1.7.21}$$

因为

$$\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^- \end{pmatrix}$$

是准对角阵

$$\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}$$

的广义逆, 将此事实应用到 (1.7.21), 再将三个矩阵相乘即得所证.

注 1　从定理的证明过程我们知道, (1.7.19) 的右端只是 A^- 的一部分, 但根据定理 1.7.4, 当 $\mathscr{M}(C) \subset \mathscr{M}(A), \mathscr{M}(B') \subset \mathscr{M}(A')$ 时, 矩阵 BA^-C 与 A^- 选择无关. 因此, 在许多重要应用场合, 我们感兴趣的只是求到一个 A^-, 而不必求出全部 A^-. 从这个意义讲, 本定理的结论仍然是十分有用的.

更进一步, 当 $A \geqslant 0$ 时, 我们可以不要求 A_{11}^{-1} 和 A_{22}^{-1} 存在, 类似的分块广义逆的公式仍然成立, 为了证明相应的结果, 我们先证明下面的引理.

引理 1.7.1　设 A 分块为 (1.7.13) 且 $A \geqslant 0$, 则矩阵 $A_{21}A_{11}^-A_{12}$, $A_{12}A_{22}^-A_{21}$ 与 A_{11}^-, A_{22}^- 的选择无关.

证明 因 $A \geqslant 0$, 依定理 1.4.2(3) 知, 存在 $B=(B_1 \vdots B_2)$, 使 $A=B^*B$, 即

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} B_1^*B_1 & B_1^*B_2 \\ B_2^*B_1 & B_2^*B_2 \end{pmatrix}. \tag{1.7.22}$$

于是

$$A_{21}A_{11}^{-}A_{12} = B_2^*B_1(B_1^*B_1)^{-}B_1^*B_2, \tag{1.7.23}$$

$$A_{12}A_{22}^{-}A_{21} = B_1^*B_2(B_2^*B_2)^{-}B_2^*B_1. \tag{1.7.24}$$

根据定理 1.7.5(1), (1.7.23) 和 (1.7.24) 之右端与所含广义逆的选择无关, 引理证毕.

定理 1.7.13 设 A 分块为 (1.7.13) 且 $A \geqslant 0$, 则

$$A^{-} = \begin{pmatrix} A_{11}^{-} + A_{11}^{-}A_{12}A_{22.1}^{-}A_{21}A_{11}^{-} & -A_{11}^{-}A_{12}A_{22.1}^{-} \\ -A_{22.1}^{-}A_{21}A_{11}^{-} & A_{22.1}^{-} \end{pmatrix}. \tag{1.7.25}$$

等价地

$$A^{-} = \begin{pmatrix} A_{11.2}^{-} & -A_{11.2}^{-}A_{12}A_{22}^{-} \\ -A_{22}^{-}A_{21}A_{11.2}^{-} & A_{22}^{-} + A_{22}^{-}A_{21}A_{11.2}^{-}A_{12}A_{22}^{-} \end{pmatrix}, \tag{1.7.26}$$

这里 $A_{22.1}=A_{22}-A_{21}A_{11}^{-}A_{12}, A_{11.2}=A_{11}-A_{12}A_{22}^{-}A_{21}$.

证明 根据引理 1.7.1 的证明, A 有表达式 (1.7.22). 应用定理 1.7.5(2), 有

$$A_{21}A_{11}^{-}A_{11} = B_2^*B_1(B_1^*B_1)^{-}B_1^*B_1 = B_2^*B_1 = A_{21}, \tag{1.7.27}$$

$$A_{11}A_{11}^{-}A_{12} = B_1^*B_1(B_1^*B_1)^{-}B_1^*B_2 = B_1^*B_2 = A_{12}. \tag{1.7.28}$$

据此, 可以验证

$$\begin{pmatrix} I & 0 \\ A_{21}A_{11}^{-} & I \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} I & -A_{11}^{-}A_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}. \tag{1.7.29}$$

此式与 (1.7.18) 类似, 应用与前面完全相同的方法可证得 (1.7.25), 用相似的方法可证明 (1.7.26).

注 2 一般说来, 对A^+, 并没有类似于定理 1.7.12 的结果. 但在条件 $r(A) = r(A_{11})+r(A_{22})$ 下, 我们有下面类似于 (1.7.25) 和 (1.7.26) 的表达式

$$\begin{aligned} A^{+} &= \begin{pmatrix} A_{11}^{+} + A_{11}^{+}A_{12}A_{22.1}^{+}A_{21}A_{11}^{+} & -A_{11}^{+}A_{12}A_{22.1}^{+} \\ -A_{22.1}^{+}A_{21}A_{11}^{+} & A_{22.1}^{+} \end{pmatrix} \\ &= \begin{pmatrix} A_{11.2}^{+} & -A_{11.2}^{+}A_{12}A_{22}^{+} \\ -A_{22}^{+}A_{21}A_{11.2}^{+} & A_{22}^{+} + A_{22}^{+}A_{21}A_{11.2}^{+}A_{12}A_{22}^{+} \end{pmatrix}. \end{aligned}$$

证明可以在文献 (Marsaglia, Styan 1974a) 中找到.

在结束这一节的时候, 我们用广义逆来表示一般相容线性方程组的解.

定理 1.7.14 设 $Ax = b$ 为相容线性方程组, $b \neq 0$, 则

(1) x 为方程组的解 $\Longleftrightarrow x = A^- b$, 其中 A^- 为 A 的任一广义逆;

(2) $x_0 = A^+ b$ 为 $Ax = b$ 的解集中长度最短者.

证明 (1) 我们先证明, 对一切 $A^-, x = A^- b$ 为一个解. 事实上, 因 $Ax = b$ 是相容的, 故存在 x_0, 使得 $Ax_0 = b$. 于是 $Ax = A(A^- b) = AA^- Ax_0 = Ax_0 = b$.

反过来, 对任一解 x_0, 我们将证明, 可找到一个 A^-, 使得 $x_0 = A^- b$. 设 G 为 A 的任一广义逆, 则容易验证, 齐次线性方程组 $Ax = 0$ 的通解可表为

$$x = (I - GA)t, \tag{1.7.30}$$

这里 t 为任意向量. 于是 $Ax = b$ 的通解可以表为

$$x = Gb + (I - GA)t, \tag{1.7.31}$$

对于现在的 x_0, 一定存在 t_0, 使得

$$x_0 = Gb + (I - GA)t_0. \tag{1.7.32}$$

根据假设 $b \neq 0$, 存在矩阵 U, 使得 $t_0 = Ub$. 事实上, U 的一个选择是 $U = t_0(b^* b)^{-1} b^*$, 代入 (1.7.32), 得

$$x_0 = Gb + (I - GA)Ub = [G + (I - GA)U]b.$$

容易直接验证, $G + (I - GA)U$ 就是一个 A^-. 得证.

(2) 在 (1.7.31) 中, 取 $G = A^+$. 这样, $Ax = b$ 的通解可表为

$$x = A^+ b + (I - A^+ A)t. \tag{1.7.33}$$

于是

$$\begin{aligned}
\|x\|^2 &= (A^+ b + (I - A^+ A)t)^* (A^+ b + (I - A^+ A)t) \\
&= \|x_0\|^2 + t^*(I - A^+ A)^2 t + 2b^*(A^+)^*(I - A^+ A)t \\
&= \|x_0\|^2 + t^*(I - A^+ A)^2 t \\
&\geqslant \|x_0\|^2.
\end{aligned}$$

等号成立 $\Longleftrightarrow (I - A^+ A)t = 0 \Longleftrightarrow x = A^+ b = x$. 定理证毕.

§1.8 幂等阵与正交投影阵

定义 1.8.1 若 n 阶方阵 A 满足 $A^2=A$, 则称 A 为幂等阵.

根据定义, 立即可以推出如下事实.

定理 1.8.1 对任意矩阵 A, A^-A, AA^-, $I-A^-A$ 和 $I-AA^-$ 都是幂等阵, 特别, $A^+A, AA^+, I-A^+A$ 和 $I-AA^+$ 都是幂等阵.

下面的定理概括了幂等阵的一些重要性质.

定理 1.8.2 设 A 为 $n\times n$ 幂等阵, 则

(1) A^* 和 $I-A$ 都是幂等阵;

(2) A 的 Jordan 标准形为 $\begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}$;

(3) A 的特征值只能是 0 或 1;

(4) $r(A)=\mathrm{tr}A$;

(5) $\mathscr{M}(I-A)=N(A)$;

(6) $\mathscr{M}(A)=N(I-A)$.

证明 (1) 和 (2) 可直接从幂等阵的定义推出. (3) 和 (4) 是 (2) 的直接推论. 容易看出, 我们只需证 (5) 和 (6) 中的一个, 下面证明 (5).

假设 $x\in\mathscr{M}(I-A)$, 则存在向量 t, 使得 $x=(I-A)t$. 因为 A 是幂等阵, 于是 $Ax=A(I-A)t=0$, 故 $x\in N(A)$.

反过来, 若 $x\in N(A)$, 则 $Ax=0$, 因而 $(I-A)x=x$. 这表明, $x\in\mathscr{M}(I-A)$. 证毕.

下面我们讨论正交投影阵.

设 $x\in C^n$, S 是 C^n 的一个子空间, 将 x 分解为

$$x=y+z, \tag{1.8.1}$$

其中 $y\in S, z\in S^{\perp}$. 此时, 称 y 为 x 在 S 上的正交投影. 若 P 为 $n\times n$ 阵, 使得对任意 $x\in C^n$, (1.8.1) 中的 y 可表为 $y=Px$, 则称 P 为向子空间 S 的正交投影阵.

我们知道, 对任一子空间 S, 一定存在矩阵 A, 使得 $S=\mathscr{M}(A)$. 下面的定理给出了正交投影阵的构造.

定理 1.8.3 设 A 为任一矩阵, 记 P_A 为向 $\mathscr{M}(A)$ 的正交投影阵. 则

$$P_A=A(A^*A)^-A^*. \tag{1.8.2}$$

证明 我们首先指出, 由定理 1.7.5(1) 知, P_A 与所含的广义逆 $(A^*A)^-$ 的选择无关.

设 B 为一矩阵, 满足 $\mathscr{M}(B)=\mathscr{M}(A)^{\perp}$, 则对任一向量 $x\in C^n$, 有分解式

$$x=At_1+Bt_2,$$

这里 t_1 和 t_2 为两个适当维数的向量. 依 P_A 的定义我们有

$$P_Ax=P_AAt_1+P_ABt_2=At_1,\ \text{对一切 } t_1,t_2 \text{ 成立}.$$

这表明, P_A 满足矩阵方程

$$\begin{cases} P_AA=A, & (1.8.3)\\ P_AB=0. & (1.8.4)\end{cases}$$

由 (1.8.4) 知

$$\mathscr{M}(P_A^*)\subset\mathscr{M}(B)^{\perp}=\mathscr{M}(A).$$

于是, 存在矩阵 X, 使得

$$P_A^*=AX. \tag{1.8.5}$$

代入 (1.8.3) 得到 $X^*A^*A=A$, 即

$$(A^*A)X=A^*. \tag{1.8.6}$$

显然, 此矩阵方程是相容的, 再应用定理 1.7.14, 可以推知 (1.8.6) 的解为 $X=(A^*A)^-A^*$. 代入 (1.8.5), 即可得 $P_A=A(A^*A)^-A^*$. 定理证毕.

若一个矩阵既是 Hermite 阵, 又是幂等阵, 则称其为 Hermite 幂等阵.

定理 1.8.4　P 为正交投影阵当且仅当它是 Hermite 幂等阵.

证明　先证必要性. 设 P 为 $\mathscr{M}(A)$ 正交投影阵. 由定理 1.8.3, $P=A(A^*A)^-A^*=A(A^*A)^+A^*$. 因为 $(A^*)^+=(A^+)^*$, 所以 $P_A^*=P_A$. 另一方面. 由定理 1.7.5(2) 得 $P_A^2=A(A^*A)^+A^*A(A^*A)^+A^*=A(A^*A)^+A^*=P_A$. 必要性得证.

下证充分性. 假设 P 为 Hermite 幂等阵, 由定理 1.4.1(2) 和定理 1.8.2(2) 知, 存在酉阵 $U=(U_1\vdots U_2)$, 使得

$$\begin{aligned} A&=(U_1U_2)\begin{pmatrix} I_t & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix} U_1^*\\ U_2^*\end{pmatrix}\\ &=U_1U_1^*=U_1(U_1^*U_1)^{-1}U_1^*,\end{aligned}$$

这里利用了 $U_1^*U_1=I_t$. 定理证毕.

在一定条件下, 正交投影阵的和、差、积仍为正交投影阵, 这些结果概括在如下三个定理中.

定理 1.8.5　设 P_1 和 P_2 为两个正交投影阵, 则

(1) $P = P_1 + P_2$ 为正交投影 $\Longleftrightarrow P_1P_2 = P_2P_1 = 0$;

(2) 当 $P_1P_2 = P_2P_1 = 0$ 时, $P = P_1 + P_2$ 为向 $\mathscr{M}(P_1) \oplus \mathscr{M}(P_2)$ 上的正交投影.

证明 (1) 充分性是容易的, 下面证明必要性. 假设 P 是一个正交投影阵, 根据定理 1.8.4 知 $P^2 = P$. 于是

$$P_1P_2 + P_2P_1 = 0. \tag{1.8.7}$$

用 P_1 分别左乘和右乘 (1.8.7), 得到

$$P_1P_2 + P_1P_2P_1 = 0, \tag{1.8.8}$$

$$P_2P_1 + P_1P_2P_1 = 0. \tag{1.8.9}$$

把 (1.8.8) 和 (1.8.9) 相加, 并利用 (1.8.7), 得到

$$P_1P_2P_1 = 0. \tag{1.8.10}$$

再由 (1.8.8) 和 (1.8.9), 便得到 $P_1P_2 = P_2P_1 = 0$.

(2) 我们只需证明

$$\mathscr{M}(P) = \mathscr{M}(P_1) \oplus \mathscr{M}(P_2). \tag{1.8.11}$$

对任一 $y \in \mathscr{M}(P)$, 存在 $x \in C^n$, 使得 $y = Px$, 于是

$$y = Px = P_1x + P_2x = y_1 + y_2,$$

这里 $y_i = P_ix \in \mathscr{M}(P_i), i = 1, 2$. 从 $P_1P_2 = 0$ 可推知 $y_1 \perp y_2$. 定理证毕.

定理 1.8.6 设 P_1 和 P_2 为两个正交投影阵, 则

(1) $P = P_1P_2$ 为正交投影阵 $\Longleftrightarrow P_1P_2 = P_2P_1$;

(2) 当 $P_1P_2 = P_2P_1$ 时, $P = P_1P_2$ 为向 $(\mathscr{M})(P_1) \cap \mathscr{M}(P_2)$ 上的正交投影阵.

这个定理的证明是容易的, 此处略去.

定理 1.8.7 设 P_1 和 P_2 为两个正交投影阵, 则

(1) $P = P_1 - P_2$ 为正交投影阵 $\Longleftrightarrow P_1P_2 = P_2P_1 = P_2$;

(2) 当 $P = P_1 - P_2$ 为正交投影阵时, P 为向 $\mathscr{M}(P_1) \cap \mathscr{M}(P_2)^{\perp}$ 上的正交投影.

证明类似于定理 1.8.5, 请读者自己去完成.

§1.9 Cauchy-Schwarz 不等式

Cauchy-Schwarz 不等式是一个非常基本的不等式, 在以后的讨论中我们要常常用到它. 鉴于它的重要性以及又难于把它划归到后续的任何一章, 于是我们在这里对它作一些专门讨论.

定理 1.9.1 设 $x, y \in C^n$, 则

(1)
$$|x^*y|^2 \leqslant x^*x \cdot y^*y, \tag{1.9.1}$$
等号成立 $\Longleftrightarrow$ y 与 x 线性相关;

(2) 设 A 为 $n \times n$ Hermite 阵且 $A \geqslant 0$, 则

$$|x^*Ay|^2 \leqslant x^*Ax \cdot y^*Ay, \tag{1.9.2}$$

等号成立 $\Longleftrightarrow$ y 与 x 线性相关;

(3) 设 A 为 $n \times n$ Hermite 阵, 且 $A > 0$, 则

$$|x^*y|^2 \leqslant x^*Ax \cdot y^*A^{-1}y, \tag{1.9.3}$$

等号成立 $\Longleftrightarrow$ x 与 $A^{-1}y$ 线性相关.

证明　(1) 当 x 和 y 至少一个为零向量时, 结论显然成立. 不妨设 $x \neq 0$, 定义

$$z = y - \frac{x^*y}{\|x\|}x,$$

则 $x^*z = 0$. 于是

$$\begin{aligned} 0 \leqslant \|z\|^2 = z^*y &= \|y\|^2 - \frac{x^*y}{\|x\|^2}x^*y \\ &= \|y\|^2 - \frac{|x^*y|^2}{\|x\|^2}, \end{aligned}$$

此即

$$|x^*y|^2 \leqslant \|x\|^2\|y\|^2,$$

等号成立 $\Longleftrightarrow z = 0 \Longleftrightarrow y$ 与 x 成比例.

(2) 因为 $A \geqslant 0$, 依定理 1.4.2(3), 存在矩阵 B, 使得 $A = B^*B$. 命 $u = Bx, v = By$, 对 u 和 v 应用 (1), 便得到 (2).

(3) 因为 $A > 0$, 所以 $A^{-1/2}$ 存在, 对 $u = A^{1/2}x$ 和 $v = A^{-1/2}y$ 应用 (1), 即得欲证.

下面的定理是 (1.9.3) 的变形和推广.

定理 1.9.2　(1) 若 Hermite 阵 $A \geqslant 0$, 则

$$\sup_{\substack{y \neq 0 \\ y \in \mathscr{M}(A)}} \frac{|x^*y|^2}{y^*A^-y} = x^*Ax;$$

(2) 若 Hermite 阵 $A > 0$, 则

$$\sup_{y \neq 0} \frac{|x^*y|^2}{y^*A^{-1}y} = x^*Ax,$$

这里 A^- 表示 A 的任一广义逆.

证明 (1) 因 $y \in \mathscr{M}(A)$, 所以存在 t, 使得 $y = At$. 依定理 1.9.1(2), 有

$$|x^*y|^2 \leqslant |x^*At|^2 \leqslant x^*Ax \cdot t^*At = x^*Ax \cdot y^*A^-y.$$

定理证毕.

(2) 是 (1) 的直接推论.

注 1 在 (1) 中, $y \in \mathscr{M}(A)$ 保证了 y^*A^-y 与 A^- 的选择无关, 于是也可写作 y^*A^+y.

注 2 对一切 $y \in \mathscr{M}(A)$ 和 $y \neq 0$, 总有 $y^*A^-y \neq 0$. 事实上, 设 $y = At$, 则 $y^*A^-y = t^*At = \|A^{1/2}t\|^2$. 易见, 由 $y^*A^-y = 0$ 可推出 $A^{1/2}t = 0$, 于是 $y = At = 0$, 这与 $y \neq 0$ 矛盾.

§1.10 Hadamard 乘积与 Kronecker 乘积

设 $A = (a_{ij})$ 和 $B = (b_{ij})$ 皆为 $m \times n$ 矩阵, 则矩阵 $C = (c_{ij}) = (a_{ij}b_{ij})$, 称为 A 与 B 的 Hadamard 乘积或 Schur 乘积, 记为 $C = A \circ B$.

矩阵的 Hadamard 乘积具有下列性质:

(1) $A \circ 0 = 0 \circ A = 0$, 这里 0 表示所有元素皆为零的零阵;

(2) $A \circ ee' = A = ee' \circ A$, 其中 $e' = (1, \cdots, 1)$;

(3) $A \circ B = B \circ A$;

(4) $(A + B) \circ C = A \circ C + B \circ C$;

(5) 设 a 与 b 皆为列向量, 则

$$(aa^*) \circ (bb^*) = (a \circ b)(a \circ b)^*; \tag{1.10.1}$$

(6) 设 a, b 和 c 皆为列向量, 则

$$a^*(b \circ c) = (a \circ \bar{b})^*c, \tag{1.10.2}$$

特别, 当 $c' = e' = (1, \cdots, 1)$ 时,

$$a^*b = (a \circ \bar{b})^*e. \tag{1.10.3}$$

定理 1.10.1 (Schur 乘积定理) 设 A 和 B 为 n 阶 Hermite 阵.

(1) 若 $A \geqslant 0, B \geqslant 0$, 则 $A \circ B \geqslant 0$;

(2) 若 $A > 0$, $B > 0$, 则 $A \circ B > 0$.

证明 (1) 应用定时 1.4.2(3), 将 A 和 B 写为如下形式

$$A = \sum_{j=1}^{k} u_j u_j^*, \quad k = r(A), \tag{1.10.4}$$

$$B=\sum_{j=1}^{m}u_jv_j^*,\quad m=r(B), \tag{1.10.5}$$

其中 u_j 和 v_j 皆为 $n\times 1$ 向量. 利用前面列举的性质, 我们得到

$$A\circ B=\left(\sum_{j=1}^{k}u_ju_j^*\right)\circ\left(\sum_{j=1}^{m}v_jv_j^*\right)=\sum_{i=1}^{k}\sum_{j=1}^{m}w_{ij}w_{ij}^*, \tag{1.10.6}$$

这里 $w_{ij}=u_i\circ v_j$, 这就证明了 $A\circ B\geqslant 0$.

(2) 我们只需证明 $A\circ B$ 可逆. 下面用反证法, 假设 $A\circ B$ 为奇异阵, 则存在非零向量 x_0, 使得 $(A\circ B)x_0=0$. 在 $A>0$ 和 $B>0$ 的假设下, (1.10.4) 和 (1.10.5) 中的 $k=m=n$, 利用 (1.10.6) 得

$$x_0^*(A\circ B)x_0=\sum_{i=1}^{n}\sum_{j=1}^{n}x_0^*(w_{ij}w_{ij}^*x_0)=\sum_{i=1}^{n}\sum_{j=1}^{n}|x_0^*w_{ij}|^2=0.$$

应用 (1.10.2), 对一切 i, j, 我们有

$$|x_0^*w_{ij}|^2=|x_0^*(u_i\circ v_j)|^2=|(x_0\circ\bar{u}_i)^*v_j|^2=0.$$

这表明, 对每个 i, 有

$$(x_0\circ\bar{u}_i)\perp v_j,\ j=1,\cdots,n.$$

于是, $x_0\circ\bar{u}_i=0, i=1,\cdots,n$. 再应用 (1.10.3), 可推得 $x_\theta^*u_i=0, i=1,\cdots,n$. 因 $u_1,\cdots,u_n$ 线性无关, 故 $x_0=0$, 与假设矛盾. 定理证毕.

注 1 对本定理的结论 (1), 还有许多其他证法, 详见许以超 (1966, p.410).

从 (1.10.6) 我们得到如下推论.

推论 1.10.1 设 A 和 B 为 n 阶 Hermite 阵, 且 $A\geqslant 0, B\geqslant 0$, 则

(1) $r(A\circ B)\leqslant r(A)\cdot r(B)$;

(2) 若 $r(A)r(B)<n$, 则 $A\circ B$ 为奇异阵.

现在我们转到矩阵的另一种乘积 ——Kronecker 乘积.

设 $A=(a_{ij})$ 和 $B=(b_{ij})$ 分别为 $m\times n$ 和 $p\times q$ 矩阵, 矩阵 $C=(a_{ij}B)$ 是一个 $mp\times nq$ 矩阵, 称为 A 与 B 的 Kronecker 乘积, 记为 $C=A\otimes B$, 即

$$A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B\\ a_{21}B & a_{22}B & \cdots & a_{2n}B\\ \vdots & \vdots & & \vdots\\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B\end{pmatrix}.$$

Kronecker 乘积具有下列性质:

(1) $0 \otimes A = A \otimes 0 = 0$, 这里 0 表示零阵;

(2) $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$,

$A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$;

(3) $I_n \otimes A = \text{diag}(A, A, \cdots, A)$;

(4) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$;

(5) $(A \otimes B)' = A' \otimes B'$;

(6) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$;

(7) $(A \otimes B)^- = A^- \otimes B^-$, 这里右端 $A^- \otimes B^-$ 只是集合 $(A \otimes B)^-$ 的一部分. 特别地, $(A \otimes B)^+ = A^+ \otimes B^+$. 当 A 和 B 皆可逆时, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

定理 1.10.2 设 A 和 B 分别为 $n \times n$ 和 $m \times m$ 阵. 记它们的特征值分别为 $\lambda_1, \cdots, \lambda_n$ 和 $\mu_1, \cdots, \mu_m$. 则 $A \otimes B$ 的 mn 个特征值为 $\lambda_i \mu_j (i = 1, \cdots, n; j = 1, \cdots, m)$.

证明 根据 Jordan 分解 (见定理 1.5.6), 存在 $n \times n$ 和 $m \times m$ 可逆阵 P 和 Q, 使得 $A = P\Lambda P^{-1}$ 和 $B = Q\Delta Q^{-1}$, 这里

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & * \\ 0 & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \qquad \Delta = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & * \\ 0 & & \ddots & \\ & & & \mu_m \end{pmatrix}$$

为两个上三角阵, 于是

$$\begin{aligned} A \otimes B &= (P\Lambda P^{-1}) \otimes (Q\Delta Q^{-1}) \\ &= (P \otimes Q)(\Lambda \otimes \Delta)(P^{-1} \otimes Q^{-1}). \end{aligned}$$

这表明 $A \otimes B$ 相似于 $\Lambda \otimes \Delta$, 而后者仍为上三角阵, 它的对角元为 $\lambda_i \mu_j (i = 1, \cdots, n; j = 1, \cdots, m)$. 于是 $A \otimes B$ 的特征值为 $\lambda_i \mu_j (i = 1, \cdots, n; j = 1, \cdots, m)$. 证毕.

由此定理我们可得到如下推论.

推论 1.10.2 设 A 和 B 分别为 $n \times n$ 和 $m \times m$ 方阵. 则

(1) $\det(A \otimes B) = (\det A)^m (\det B)^n$;

(2) $\text{tr}(A \otimes B) = (\text{tr} A)(\text{tr} B)$;

(3) $r(A \otimes B) = r(A) \cdot r(B)$;

(4) 若 $A \geqslant 0, B \geqslant 0$, 则 $A \otimes B \geqslant 0$. 特别 $A > 0, B > 0$, 则 $A \otimes B > 0$.

由定理及推论 1.2.1 立即得到 (1) 和 (2), (3) 可由秩的定义直接导出, 最后, 结合定理 1.4.2 和定理 1.4.3 推出 (4).

最后我们给出 Hadamard 乘积与 Kronecker 乘积的一个关系.

定理 1.10.3　设 A 和 B 皆为 n 阶方阵. 记 $J=\{1,n+2,2n+3,3n+4,\cdots,n^2\}$, $A(J)$ 表示由 A 的子方阵 $A(1,n+2,2n+3,\cdots,n^2)$. 则

$$A\circ B=(A\otimes B)(J). \tag{1.10.7}$$

证明容易根据定义完成.

§1.11 矩阵微商

假设 $X=(x_{ij})$ 是一个 $m\times n$ 实矩阵, $u=f(X)$ 为一实值函数. 在求 $f(X)$ 的极值时, 我们总需要计算导数 $\dfrac{\partial f}{\partial x_{ij}}(i=1,\cdots,m;j=1,\cdots,n)$. 这里总共有 mn 个导数, 为简单计, 我们记

$$\frac{\partial f}{\partial X}=\begin{pmatrix}\dfrac{\partial f}{\partial x_{11}} & \cdots & \dfrac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \dfrac{\partial f}{\partial x_{m1}} & \cdots & \dfrac{\partial f}{\partial x_{mn}}\end{pmatrix}.$$

这是一个 $m\times n$ 矩阵, 称为 u 对矩阵 X 的导数. 从本质上讲, 对矩阵 X 求导数, 和对普通多元函数的求导是一样的, 但是为了表示上的简洁和运算上的方便, 我们需要把所求出的 mn 个偏导数 $\dfrac{\partial f}{\partial x_{ij}}$ 写成矩阵形式, 因而需要一些独特的运算规律和技巧. 本节我们讨论这方面的一些最基本的内容.

在这一节, 我们假定所讨论的矩阵和向量都是实的. 下面是两个最简单的例子.

例 1.11.1　设 a 和 x 均为 $n\times 1$ 向量, $u=a'x$, 则 $\dfrac{\partial u}{\partial x}=a$.

例 1.11.2　设 A 为 $n\times n$ 对称阵, x 为 $n\times 1$ 向量, $u=x'Ax$, 则 $\dfrac{\partial u}{\partial x}=2Ax$.

记 $E_{ij}(m\times n)$ 为第 (i,j) 元为 1, 其余元素皆为零的 $m\times n$ 矩阵, 在不致引起混淆时, 常简记为 E_{ij}. 显然

$$\sum_{i=1}^{m}E_{ii}(m\times m)=I_m,$$

$$\sum_{i,j}x_{ij}E_{ij}(m\times n)=X.$$

$(X)_{ij}$ 表示矩阵 X 的第 (i,j) 元 x_{ij}. 若 $Z=(z_{ij}(t))_{\mathrm{m\times n}}$, 即 Z 的所有元素都是 t 的一元函数, 则记

$$\frac{dZ}{dt}=\begin{pmatrix}\dfrac{dz_{11}}{dt} & \cdots & \dfrac{dz_{1n}}{dt}\\ \vdots & & \vdots\\ \dfrac{dz_{m1}}{dt} & \cdots & \dfrac{dz_{mn}}{dt}\end{pmatrix}.$$

在矩阵微商中, 下面的转换定理是一个重要工具.

定理 1.11.1 (转换定理) 设 $X=(x_{ij})$ 和 $Y=(y_{ij})$ 分别为 $m\times n$ 和 $p\times q$ 矩阵, A, B, C 和 D 分别为 $p\times m$, $n\times q$, $p\times n$ 和 $m\times q$ 的矩阵, 它们可以是 X 的函数, 则如下两条是等价的.

(1) $\dfrac{\partial Y}{\partial x_{ij}}=AE_{ij}(m\times n)B+CE_{ij}'(m\times n)D,\quad i=1,\cdots,m;\ j=1,\cdots,n;$

(2) $\dfrac{\partial y_{ij}}{\partial X}=A'E_{ij}(p\times q)B'+DE_{ij}'(p\times q)C,\quad i=1,\cdots,p;\ j=1,\cdots,q.$

证明 记 $e_i=(0,\cdots,1,0,\cdots,0)'$, 即 e_i 是第 i 个元素为 1, 其余元素皆为零的向量, 它的维数由上下文而定, 则 $E_{ij}(p\times q)=e_ie_j'$, 这里 e_i 为 $p\times 1$, 而 e_j 为 $q\times 1, e_i'Ae_i=a_{ij}$, 这里 e_i 为 $p\times 1$, 而 e_j 为 $m\times 1$.

因为

$$\begin{aligned}&e_k'(AE_{ij}B+CE_{ij}'D)e_l\\=&e_k'Ae_ie_j'Be_l+e_k'Ce_je_i'De_l\\=&e_i'A'e_ke_l'B'e_j+e_i'De_le_k'Ce_j\\=&e_i'(A'e_ke_l'B'+De_le_k'C)e_j\\=&e_i'(A'E_{kl}B'+DE_{lk}C)e_j.\end{aligned}$$

若 (1) 成立, 则矩阵 $\dfrac{\partial Y}{\partial x_{ij}}$ 的 (k,l) 元

$$\begin{aligned}\left(\frac{\partial Y}{\partial x_{ij}}\right)_{kl}&=e_k'(AE_{ij}B+CE_{ij}'D)e_l\\&=e_i'(A'E_{kl}B'+DE_{lk}C)e_j.\end{aligned}\tag{1.11.1}$$

但另一方面

$$\left(\frac{\partial Y}{\partial x_{ij}}\right)_{kl}=\frac{\partial y_{kl}}{\partial x_{ij}}=\left(\frac{\partial y_{kl}}{\partial X}\right)_{ij}.\tag{1.11.2}$$

由 (1.11.1) 和 (1.11.2) 我们得到

$$\left(\frac{\partial y_{kl}}{\partial X}\right)_{ij}=e_i'(A'E_{kl}B'+DE_{lk}C)e_i,$$

也就是

$$\frac{\partial y_{kl}}{\partial X}=A'E_{kl}B'+DE_{kl}'C,$$

即 (2). 用完全同样方法可从 (2) 推出 (1). 定理证毕.

迹和行列式是最常见的两类矩阵函数. 下面我们通过这两类函数的求导, 说明转换定理的应用.

例 1.11.3 设 A 为 $n\times m$ 矩阵, X 为 $m\times n$ 矩阵, 则 $\dfrac{\partial \mathrm{tr}AX}{\partial X}=A'$.

证明 $$\frac{\partial \text{tr} AX}{\partial X} = \frac{\partial}{\partial X}\left(\sum_{i=1}^{n}(AX)_{ii}\right) = \sum_{i=1}^{n}\frac{\partial (AX)_{ii}}{\partial X}, \tag{1.11.3}$$

因为

$$\frac{\partial AX}{\partial x_{ij}} = AE_{ij}(m \times n).$$

由转换定理, 得

$$\frac{\partial (AX)_{ij}}{\partial X} = A'E_{ij}(n \times n).$$

代入 (1.11.3), 则有

$$\frac{\partial \text{tr} AX}{\partial X} = \sum_{i=1}^{n} A'E_{ii}(n \times n) = A'\sum_{i=1}^{n} E_{ii}(n \times n) = A'.$$

证毕.

在这个例子中, 我们本来需要求 $\dfrac{\partial (AX)_{ii}}{\partial x}$, 但它计算很困难. 应用转换定理我们把问题归结为求 $\dfrac{\partial AX}{\partial x_{ij}}$, 这个很容易求到.

例 1.11.4　设 A, X 和 B 分别为 $m\times n$, $n\times k$ 和 $k\times m$ 矩阵, 则 $\dfrac{\partial \text{tr} AXB}{\partial X} = A'B'$.

证明 $$\frac{\partial \text{tr} AXB}{\partial X} = \frac{\partial}{\partial X}\left(\sum_{i=1}^{m}(AXB)_{ii}\right) = \sum_{i=1}^{m}\frac{\partial (AXB)_{ii}}{\partial X}, \tag{1.11.4}$$

因为

$$\frac{\partial AXB}{\partial x_{ij}} = AE_{ij}(n \times k)B,$$

依转换定理

$$\frac{\partial (AXB)_{ij}}{\partial X} = A'E_{ij}(m \times m)B'.$$

代入 (1.11.4) 得

$$\frac{\partial \text{tr} AXB}{\partial X} = \sum_{i=1}^{m} A'E_{ii}(m \times m)B' = A'B'.$$

证毕.

例 1.11.5　(1) 设 A 为 $n\times n$ 方阵, X 为 $n\times m$ 阵, 则 $\dfrac{\partial \text{tr} X'AX}{\partial X} = (A'+A)X$;

(2) 设 A 为 $m\times m$ 方阵, X 为 $n\times m$ 阵, 则 $\dfrac{\partial \text{tr} XAX'}{\partial X} = X(A+A')$.

证明　(2) 和 (1) 的证明类似, 以下只证 (1). 首先

$$\frac{\partial \text{tr}(X'AX)}{\partial X} = \sum_{i=1}^{m}\frac{\partial (X'AX)_{ii}}{\partial X}. \tag{1.11.5}$$

因为

$$\frac{\partial X'AX}{\partial x_{ij}} E'_{ij}AX + X'AE_{ij},$$

利用转换定理, 得

$$\frac{\partial (X'AX)_{ii}}{\partial X} = A'XE_{ii} + AXE'_{ii}.$$

代入 (1.11.5), 我们得到

$$\frac{\partial \mathrm{tr}(X'AX)}{\partial X} = \sum_{i=1}^{m} (A'XE_{ii} + AXE'_{ii}) = (A' + A)X.$$

证毕.

为了导出另一个例子, 我们先证如下引理.

引理 1.11.1 设 $A(x)$ 为 $n \times n$ 可逆阵, 其元素为单变量 x 的函数, $A^{-1}(x)$ 表示 $A(x)$ 的逆矩阵, 则

$$\frac{\mathrm{d}A^{-1}(x)}{\mathrm{d}x} = -A^{-1}(x)\frac{\mathrm{d}A(x)}{\mathrm{d}x}A^{-1}(x). \tag{1.11.6}$$

证明 对矩阵恒等式 $A^{-1}(x)A(x) = I_n$ 两边求导数, 我们得

$$\frac{\mathrm{d}A^{-1}(x)}{\mathrm{d}x} \cdot A(x) + A^{-1}(x)\frac{\mathrm{d}A(x)}{\mathrm{d}x} = 0,$$

由此立得 (1.11.6). 证毕.

例 1.11.6 设 A 为 $n \times n$ 方阵, X 为 $n \times m$ 矩阵, $X'AX$ 可逆. 则

$$\frac{\partial \mathrm{tr}(X'AX)^{-1}}{\partial X} = -2AX(X'AX)^{-2}.$$

证明 $$\frac{\partial \mathrm{tr}(X'AX)^{-1}}{\partial X} = \sum_{i=1}^{m} \frac{\partial ((X'AX)^{-1})_{ii}}{\partial X}. \tag{1.11.7}$$

利用引理 1.11.1, 有

$$\begin{aligned}
\frac{\partial (X'AX)^{-1}}{\partial x_{ij}} &= -(X'AX)^{-1}\frac{\partial X'AX}{\partial x_{ij}}(X'AX)^{-1} \\
&= -(X'AX)^{-1}[E'_{ij}(n \times m)AX + X'AE_{ij}(n \times m)](X'AX)^{-1} \\
&= -(X'AX)^{-1}E'_{ij}(n \times m)AX(X'AX)^{-1} \\
&\quad -(X'AX)^{-1}X'AE_{ij}(n \times m)(X'AX)^{-1}.
\end{aligned}$$

由转换定理

$$\begin{aligned}
\frac{\partial ((X'AX)^{-1})_{ii}}{\partial X} = &-A'X(X'AX)^{-1}E_{ii}(m \times m)(X'AX)^{-1} \\
&-AX(X'AX)^{-1}E'_{ii}(m \times m)(X'AX)^{-1},
\end{aligned}$$

代入 (1.11.7), 得

$$\begin{aligned}\frac{\partial \text{tr}(X'AX)^{-1}}{\partial X} &= -A'X(X'AX)^{-1}\sum_{i=1}^{m}E_{ii}(m\times m)(X'AX)^{-1}\\&\quad -A'X(X'AX)^{-1}\sum_{i=1}^{m}E'_{ii}(m\times m)(X'AX)^{-1}\\&= -2A'X(X'AX)^{-2}.\end{aligned}$$

证毕.

下面我们证明另一个引理, 它在涉及行列式的求导中很有用.

引理 1.11.2　设 $u=u(Y)$, 这里 $Y=(y_{ij})$ 为 $m\times n$ 矩阵, 而 Y 的元素 y_{ij} 又是矩阵 X 的函数, 则

$$\frac{\partial u}{\partial X}=\sum_{ij}\left(\frac{\partial u}{\partial Y}\right)_{ij}\frac{\partial y_{ij}}{\partial X}, \tag{1.11.8}$$

其中 $\left(\dfrac{\partial u}{\partial Y}\right)_{ij}$ 表示矩阵 $\dfrac{\partial u}{\partial Y}$ 的 (i,j) 元, 根据复合函数的链式求导法则, 立得

$$\frac{\partial u}{\partial X}=\left(\frac{\partial u}{\partial x_{kl}}\right)=\left(\sum_{ij}\frac{\partial u}{\partial y_{ij}}\frac{\partial y_{ij}}{\partial x_{kl}}\right)=\sum_{ij}\left(\frac{\partial u}{\partial Y}\right)_{ij}\frac{\partial y_{ij}}{\partial X}.$$

例 1.11.7　设 X 为 $n\times n$ 可逆方阵, 则

$$\frac{\partial \text{det}X}{\partial X}=\text{det}X\cdot(X^{-1})'.$$

证明　由行列式的 Laplace 展开 (见定理 3.1.1),

$$\text{det}X=\sum_{j=1}^{n}x_{ij}A_{ij},$$

这里 A_{ij} 为 X 的 (i,j) 元的代数余子式, 若用 M_{ij} 表示将 X 的第 i 行和第 j 列删去后剩下的子阵的行列式, 则 $A_{ij}=(-1)^{i+j}M_{ij}$, 因为 A_{ij} 与 x_{ij} 无关, 于是

$$\frac{\partial \text{det}X}{\partial X}=\left(\frac{\partial \text{det}X}{\partial x_{ij}}\right)=(A_{ij})=\text{det}X(X^{-1})'.$$

证毕.

例 1.11.8　设 A 为 $n\times n$ 方阵, X 为 $n\times m$ 矩阵, 且 $X'AX$ 可逆. 则

$$\frac{\partial \text{lndet}X'AX}{\partial X}=(A'+A)X(X'AX)^{-1}.$$

证明 记 $Y = X'AX$. 由引理 1.11.2 和例 1.11.7, 得

$$
\begin{aligned}
\frac{\partial \ln\det X'AX}{\partial X} &= \frac{1}{\det Y}\frac{\partial \det Y}{\partial X} \\
&= \frac{1}{\det Y}\sum_{ij}\left(\frac{\partial \det Y}{\partial Y}\right)_{ij}\frac{\partial y_{ij}}{\partial X} \\
&= \frac{1}{\det Y}\sum_{ij}\det Y\cdot (Y^{-1})'_{ij}\frac{\partial y_{ij}}{\partial X} \\
&= \sum_{ij}(X'AX)^{-1}_{ij}\frac{\partial (X'AX)_{ij}}{\partial X}.
\end{aligned} \tag{1.11.9}
$$

因为

$$\frac{\partial X'AX}{\partial x_{ij}} = E'_{ij}AX + X'AE_{ij},$$

由转换定理, 得

$$\frac{\partial (X'AX)_{ij}}{\partial X} = A'XE_{ij} + AXE'_{ij}.$$

代入 (1.11.9), 我们有

$$
\begin{aligned}
\frac{\partial \ln\det X'AX}{\partial X} &= \sum_{ij}(X'AX)^{-1}_{ij}(A'XE_{ij} + AXE'_{ij}) \\
&= A'X\sum_{ij}(X'AX)^{-1}_{ij}E_{ij} + AX\sum_{ij}(X'AX)^{-1}_{ij}E'_{ij} \\
&= A'X(X'AX)^{-1} + AX(X'AX)^{-1} \\
&= (A'+A)X(X'AX)^{-1}.
\end{aligned}
$$

证毕.

例 1.11.9 设 A, X 和 B 分别为 $m\times n, n\times k$ 和 $k\times m$ 矩阵, 且 AXB 可逆, 则

$$\frac{\partial \ln\det AXB}{\partial X} = A'(B'X'A')^{-1}B'.$$

证明方法类似于例 1.11.8, 留给读者作为练习.

第 2 章　秩

矩阵秩的概念是由 Sylvester 于 1861 年引进的 (Mirsky 1955, p.136), 它是矩阵的最重要的数字特征之一. 本章我们将讨论有关矩阵秩的一些重要不等式.

§2.1 给出了矩阵秩的一些基本性质; §2.2 和 §2.3 证明了两个和三个矩阵积的秩的不等式, 即 Sylvester 定律和 Frobenius 不等式; §2.4 讨论了两矩阵和的秩的不等式; 在本章最后一节, 我们给出了估计一个矩阵秩的下界的两个不等式.

§2.1　基 本 性 质

定义 2.1.1　矩阵 A 的秩定义为它的行 (或列) 向量的极大线性无关组所包含的向量个数, 记为 $r(A)$. 若 A 为 $m\times n$ 矩阵, 当 $r(A)=m$ 时, 称 A 为行满秩; 当 $r(A)=n$ 时, 则称 A 为列满秩; 当 $r(A)=m=n$ 时, 称 A 为满秩阵或可逆阵.

定理 2.1.1　(1) $r(AB)\leqslant \min(r(A),r(B))$;

(2) $r(A+B)\leqslant r(A\vdots B)\leqslant r(A)+r(B)$.

证明　(1) 因为 $\mathscr{M}(AB)\subset\mathscr{M}(A),\mathscr{M}(B'A')\subset\mathscr{M}(B')$, 应用 (1.1.1), 我们有

$$\begin{aligned}r(AB)&=\dim\mathscr{M}(AB)\leqslant\dim\mathscr{M}(A)=r(A),\\ r(B'A')&=\dim\mathscr{M}(B'A')\leqslant\dim\mathscr{M}(B')=r(B').\end{aligned}$$

再由 $r(C)=r(C')$ 对一切矩阵 C 成立, 结论得证.

(2) 设 $r(A)=p,r(B)=q$, 且 $a_1,\cdots,a_p$ 和 $b_1,\cdots,b_q$ 分别为 $\mathscr{M}(A)$ 和 $\mathscr{M}(B)$ 的基. 则 $\mathscr{M}(A\vdots B)$ 中的任一向量均可由 $a_1,\cdots,a_p,b_1,\cdots,b_q$ 线性表示. 于是

$$r(A\vdots B)=\dim\mathscr{M}(A\vdots B)\leqslant p+q.$$

右边不等式得证.

假设 A 和 B 皆为 $m\times n$ 矩阵. 则

$$A+B=(A\vdots B)\begin{pmatrix}I_n\\ I_n\end{pmatrix},$$

于是 $\mathscr{M}(A+B)\subset\mathscr{M}(A\vdots B)$. 再应用 (1.1.1), 得

$$r(A+B)=\dim\mathscr{M}(A+B)\leqslant\dim\mathscr{M}(A\vdots B)=r(A\vdots B).$$

定理证毕.

定理 2.1.2 $r(AA^*)=r(A)$.

证明 事实上, 我们能够证明如下更强的结论

$$\mathscr{M}(AA^*)=\mathscr{M}(A). \tag{2.1.1}$$

显然, $\mathscr{M}(AA^*)\subset\mathscr{M}(A)$, 故只需证

$$\mathscr{M}(A)\subset\mathscr{M}(AA^*). \tag{2.1.2}$$

设 $u\perp\mathscr{M}(AA^*)$, 则 $u^*AA^*=0$, 因而 $u^*AA^*u=\|A^*u\|^2=0$. 这表明 $A^*u=0$, 此即 $u\perp\mathscr{M}(A)$. 这就证明了 (2.1.2). 定理证毕.

定理 2.1.3 (消去法则) 设 A 为 $m\times n$ 且秩为 r 的矩阵, X 为 $p\times m$ 的列满秩阵, 而 Y 为 $n\times q$ 的行满秩阵, 则

$$r(A)=r(XA)=r(AY)=r(XAY).$$

证明 因 X 为列满秩, 定义 $Q=(X'X)^{-1}X'$, 则有 $QX=I_m$. 应用定理 2.1.1.(1), 得

$$r(A)=r(QXA)\leqslant r(XA)\leqslant r(A),$$

于是 $r(A)=r(XA)$. 类似地, 可以证明 $r(A)=r(AY)$. 将 AY 视为 A, 应用第一个等式, 即得 $r(A)=r(XAY)$. 定理证毕.

推论 2.1.1 设 A 为 $m\times n$ 阵, P 和 Q 为任意的 $m\times m$ 和 $n\times n$ 可逆阵, 则

$$r(A)=r(PA)=r(AQ)=r(PAQ).$$

§2.2 Sylvester 定律

关于两个矩阵积的秩, 最重要的不等式是所谓的Sylvester 定律. 它是由 Sylvester 于 1884 年首先证明的 (Marsaglia, Styan 1974b), 后来人们给出了很多种证明. Marsaglia 和 Styan(1974b) 获得了等号成立的一些充要条件.

定理 2.2.1 (Sylvester 定律) 设 A 和 B 分别为 $m\times n$ 和 $n\times l$ 矩阵, 则

$$r(A)+r(B)-n\leqslant r(AB)\leqslant\min(r(A),r(B)). \tag{2.2.1}$$

证明 右边不等式已在定理 2.1.1 证明过. 下面证左边不等式. 对矩阵 A 应用定理 1.5.2, 存在可逆阵 P 和 Q, 使得

$$PAQ=\begin{pmatrix} I_t & 0\\ 0 & 0\end{pmatrix},$$

这里 $t=r(A)$. 再应用消去法则 (定理 2.13), 得

$$\begin{aligned} r(AB) &= r(PAQQ^{-1}B) \\ &= r\left(\begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix} D\right), \end{aligned} \tag{2.2.2}$$

其中 $\underset{n\times l}{D}=(d_{ij})=Q^{-1}B$. 将其分块为

$$D=\begin{pmatrix} D_1 \\ D_2 \end{pmatrix}, \quad \text{其中 } D_1 \text{ 为 } t\times l, D_2 \text{ 为 } (n-t)\times l.$$

于是, (2.2.2) 变形为

$$r(AB)=r(D_1), \tag{2.2.3}$$

问题归结为求 $r(D_1)$ 的最小值.

因为 Q 为可逆阵, 所以 $r(D)=r(B)$. 若记 $q=r(B)$, 则 D 的 n 个行向量的极大线性无关组含有 q 个向量, 当这 q 个线性无关的行位于 D 的最后 q 行时, $r(D_1)$ 达到最小, 此时 D_1 的 t 个行中至多含有 $n-q$ 个线性相关的行向量. 也就是说, D_1 的 t 个行中至少含有 $t-(n-q)$ 个线性无关的向量, 即 $r(D_1)\geqslant t+q-n=r(A)+r(B)-n$. 结合 (2.2.3), 定理得证.

接下来, 我们建立 (2.2.1) 式等号成立的充要条件. 为此, 先证明几个引理.

引理 2.2.1 对任意的 A^- 及可以相乘的 B 和 C.

(1) $r(AB,(I-AA^-)C)=r(AB)+r((I-AA^-)C)$; (2.2.4)

(2) $r\begin{pmatrix} BA \\ C(I-A^-A) \end{pmatrix}=r(BA)+r(C(I-A^-A))$. (2.2.5)

证明 (1) 不难看出, 我们只需证明

$$\mathscr{M}(AB)\cap\mathscr{M}((I-AA^-)C)=\{0\}. \tag{2.2.6}$$

假设 x_0 属于等式左边. 则存在向量 t_1 和 t_2, 使得

$$x_0=ABt_1=(I-AA^-)Ct_2.$$

用 $I-AA^-$ 左乘上式, 得 $(I-AA^-)Ct_2=0$, 即 $x_0=0$, 这就证明了 (2.2.6).

(2) 的证明与 (1) 类似. 从略.

引理 2.2.2 对任意的 A^- 和 B^-, 有

$$\begin{aligned} (1)\ r(A\vdots B) &= r(A)+r((I-AA^-)B) \\ &= r((I-BB^-)A)+r(B); \end{aligned} \tag{2.2.7}$$

(2)
$$r\begin{pmatrix} A \\ B \end{pmatrix} = r(A) + r(B(I - A^- A))$$
$$= r(A(I - B^- B)) + r(B). \tag{2.2.8}$$

证明 应用消去法则 (定理 2.1.3) 和引理 2.2.1, 我们有

$$r(A\vdots B) = r\left((A\vdots B)\begin{pmatrix} I & -A^- B \\ 0 & I \end{pmatrix}\right) = r(A\vdots(I - AA^-)B)$$
$$= r(A) + r(I - AA^-)B).$$

类似地, 可证 (2.2.7) 第二等式及 (2.2.8). 证毕.

引理 2.2.3 设 A 有 n 列, B 有 n 行, 那么对任意的 A^- 和 B^-,

$$r\begin{pmatrix} 0 & A \\ B & I_n \end{pmatrix} = r(A) + r(B\vdots I - A^- A) \tag{2.2.9}$$
$$= r\begin{pmatrix} A \\ I - BB^- \end{pmatrix} + r(B) \tag{2.2.10}$$
$$= r(A) + r(B) + r((I - BB^-)(I - A^- A) \tag{2.2.11}$$
$$= n + r(AB). \tag{2.2.12}$$

证明 应用 (2.2.8), 得

$$r\begin{pmatrix} 0 & A \\ B & I_n \end{pmatrix} = r(0\vdots A) + r(B\vdots I)(I - (0\vdots A)^-(0\vdots A)). \tag{2.2.13}$$

因为 $\begin{pmatrix} 0 \\ A^- \end{pmatrix}$ 是 $(0\vdots A)$ 的一个广义逆, 于是

$$I - (0\vdots A)^-(0\vdots A) = I - \begin{pmatrix} 0 \\ A^- \end{pmatrix}(0\vdots A) = \begin{pmatrix} I & 0 \\ 0 & I - A^- A \end{pmatrix},$$

代入 (2.2.13) 即得 (2.2.9). 类似地, 可以证明 (2.2.10).

应用 (2.2.7), 有

$$r(B\vdots I - A^- A) = r(B) + r(I - BB^-)(I - A^- A),$$

结合 (2.2.9), 得到 (2.2.11). 最后用矩阵恒等式

$$\begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}\begin{pmatrix} 0 & A \\ B & I_n \end{pmatrix}\begin{pmatrix} I & 0 \\ -B & I \end{pmatrix} = \begin{pmatrix} -AB & 0 \\ 0 & I \end{pmatrix}$$

及定理 2.1.3(消去法则), 可得 (2.2.12). 引理证毕.

下面的定理建立了 Sylvester 定律中等式成立的充要条件.

定理 2.2.2　在 Sylvester 定律中, 对任意选定的 A^- 和 B^-, 有

(1) $r(AB)=r(A)+r(B)-n \Longleftrightarrow (I-BB^-)(I-A^-A)=0$;

(2) $r(AB)=r(A) \Longleftrightarrow (B \vdots I-A^-A)$ 为行满秩;

(3) $r(AB)=r(B) \Longleftrightarrow \begin{pmatrix} A \\ I-BB^- \end{pmatrix}$ 为列满秩.

由 (2.2.11) 和 (2.2.12) 直接推得 (1); 由 (2.2.9) 和 (2.2.12) 推得 (2); 最后, 由 (2.2.10) 和 (2.2.12) 推出 (3).

注　在本定理中, 结论成立与否, 跟选定的广义逆无关. 也就是说对某一个 A^- 与 B^-, $(I-BB^-)(I-A^-A)=0$. 则对一切其他 A^-, B^-, 该等式仍成立.

§2.3　Frobenius 不等式

上节的 Sylvester 定律是关于两个矩阵乘积的秩的不等式, 这一节我们要讨论的是关于三个矩阵积的秩的不等式, 它是由 Frobenius 于 1961 年证明的 (Marsaglia 和 Styan　1974b).

我们先给出一个引理.

引理 2.3.1　对任意的 $(AB)^-$ 和 $(BC)^-$

$$r\begin{pmatrix} 0 & AB \\ BC & B \end{pmatrix} = r(AB)+r(BC)+r((I-(BC)(BC)^-)B(I-(AB)^-(AB)) \tag{2.3.1}$$

$$= r(B)+r(ABC). \tag{2.3.2}$$

证明　应用与上节 (2.2.11) 证明完全同样的方法可证 (2.3.1). 而 (2.3.2) 由消去法则 (见定理 2.1.3) 及下面的恒等式导出

$$\begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}\begin{pmatrix} 0 & AB \\ BC & B \end{pmatrix}\begin{pmatrix} I & 0 \\ -C & I \end{pmatrix}=\begin{pmatrix} -ABC & 0 \\ 0 & B \end{pmatrix}.$$

由引理 2.3.1, 立即可得到下面的定理.

定理 2.3.1 (Frobenius)

$$r(ABC) \geqslant r(AB)+r(BC)-r(B), \tag{2.3.3}$$

且等号成立当且仅当对任意选定的 $(AB)^-$ 和 $(BC)^-$,

$$(I-(BC)BC)^-)B(I-(AB)^-(AB))=0, \tag{2.3.4}$$

注 1 由 (2.3.1) 知, (2.3.4) 成立与广义逆 $(AB)^-$ 和 $(BC)^-$ 的选择无关. 也就是说, 若对某两个广义逆 $(AB)^-$ 和 $(BC)^-$, (2.3.4) 成立, 则对所有的广义逆 $(AB)^-$ 和 $(BC)^-$, (2.3.4) 也成立. 因此, 定理 2.3.1 的后一半结论可叙述为: 若 (2.3.3) 中等号成立, 则对一切 $(AB)^-$ 和 $(BC)^-$, (2.3.4) 成立. 反过来, 若对某两个特定的 $(AB)^-$ 和 $(BC)^-$, (2.3.4) 成立, 则 (2.3.3) 等号成立.

注 2 若不考虑 (2.3.3) 等号成立的条件, 不等式 (2.3.3) 有多种证明. 下面给出一个较简单的证法.

假设 B 为 $n\times l$ 矩阵, 且 $r(B)=t$. 将 B 作满秩分解 (见定理 1.5.3), $B=D_1D_2$, 这里 D_1 和 D_2 分别为 $n\times t$ 和 $t\times l$ 矩阵, 且 $r(D_1)=r(D_2)=t$. 应用 Sylvester 定律, 有

$$\begin{aligned} r(ABC) &= r(AD_1\cdot D_2C) \\ &\geqslant r(AD_1)+r(D_2C)-t \\ &\geqslant r(AD_1D_2)+r(D_1D_2C)-r(B) \\ &= r(AB)+r(BC)-r(B). \end{aligned}$$

§2.4 矩阵和的秩

对于任意两个同阶矩阵 A 和 B, 定理 2.1.1 证明了

$$r(A+B)\leqslant (A\vdots B)\leqslant r(A)+r(B).$$

本节除了改进 $r(A+B)$ 的上界外, 还将导出它的下界. 为此, 先给出如下引理.

引理 2.4.1 设 A 和 B 为同阶矩阵, 则对 $(A\vdots B)^-$ 和 $\begin{pmatrix} A \\ B \end{pmatrix}^-$ 的任一选择,

$$r\begin{pmatrix} O & A & B \\ A & A & O \\ B & O & B \end{pmatrix} = r(A\vdots B)+r\begin{pmatrix} A \\ B \end{pmatrix}+r(Q) \tag{2.4.1}$$

$$= r(A)+r(B)+r(A+B), \tag{2.4.2}$$

其中

$$Q=\left(I-\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-\right)\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}(I-(A\vdots B)^-(A\vdots B)). \tag{2.4.3}$$

证明 在 (2.3.1) 中用

$$(A\vdots B), \begin{pmatrix} A \\ B \end{pmatrix}, \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

分别代替 AB, BC 和 B, 便得到 (2.4.1) 和 (2.4.2).

注　设 A 和 B 为 $m\times n$ 矩阵, 这相当于用

$$(I_m\vdots I_m),\begin{pmatrix} A & 0\\ 0 & B\end{pmatrix},\begin{pmatrix} I_n\\ I_n\end{pmatrix}$$

分别代替 (2.3.1) 中的 A, B 和 C. 证毕.

定理 2.4.1　设 A, B 为同阶矩阵, 则

$$\begin{aligned} & r(A\vdots B)+r\begin{pmatrix} A\\ B\end{pmatrix}-r(A)-r(B)\\ & \leqslant r(A+B)\leqslant \min\left(r(A\vdots B), r\begin{pmatrix} A\\ B\end{pmatrix}\right); \end{aligned} \tag{2.4.4}$$

$$r(A)+r(B)-c-d\leqslant r(A+B)\leqslant r(A)+r(B)-\max(c,d); \tag{2.4.5}$$

$$r(Q)\leqslant \min(c,d), \tag{2.4.6}$$

这里 Q 由 (2.4.3) 所定义,

$$\begin{aligned} c&=\dim(\mathscr{M}(A)\cap\mathscr{M}(B)),\\ d&=\dim(\mathscr{M}(A')\cap\mathscr{M}(B')), \end{aligned}$$

并且, (2.4.4) 或 (2.4.5) 左边等号成立 $\Longleftrightarrow Q=0$, 而它们任一个的右边等号成立 $\Longleftrightarrow r(Q)=\min(c,d)$.

证明　由 (2.4.1) 和 (2.4.2) 得

$$r(A+B)=r(A\vdots B)+r\begin{pmatrix} A\\ B\end{pmatrix}-r(A)-r(B)+r(Q). \tag{2.4.7}$$

由此立得 (2.4.4) 左边. 再应用定理 1.1.1, 得

$$\begin{aligned} r(A\vdots B)&=\dim\mathscr{M}(A\vdots B)=\dim(\mathscr{M}(A)+\mathscr{M}(B))\\ &=\dim\mathscr{M}(A)+\dim\mathscr{M}(B)-\dim(\mathscr{M}(A)\cap\mathscr{M}(B))\\ &=r(A)+r(B)-c. \end{aligned} \tag{2.4.8}$$

类似地, 可以证明

$$r\begin{pmatrix} A\\ B\end{pmatrix}=r(A)+r(B)-d. \tag{2.4.9}$$

综合 (2.4.7), (2.4.8) 和 (2.4.9) 得到

$$r(A+B)=r(A)+r(B)-c-d+r(Q), \tag{2.4.10}$$

于是 (2.4.5) 的左边得证. 从 (2.4.7) 和 (2.4.10) 可以看到, (2.4.4) 或 (2.4.5) 左边等号成立 $\Longleftrightarrow Q=0$.

设 A, B 皆为 $m\times n$ 阵, 则

$$A+B=(A\vdots B)\begin{pmatrix} I_n \\ I_n \end{pmatrix}=(I_m\vdots I_m)\begin{pmatrix} A \\ B \end{pmatrix},$$

应用定理 2.1.1(1), 便得到 (2.4.4) 之右边.

因为

$$\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^- \text{ 和 } I-\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-$$

都是幂等阵, 应用定理 2.1.1(1) 和定理 1.8.2(4), 我们有

$$\begin{aligned} r(Q) &\leqslant r\left(I_{2m}-\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-\right) \\ &= \operatorname{tr}\left(I_{2m}-\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-\right) \\ &= 2m-\operatorname{tr}\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^- \\ &= 2m-r\left(\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-\right). \end{aligned} \tag{2.4.11}$$

再应用定理 1.7.3(5), 得

$$r\left(\begin{pmatrix} A \\ B \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}^-\right)=r\begin{pmatrix} A \\ B \end{pmatrix}. \tag{2.4.12}$$

将 (2.4.12) 代入 (2.4.11), 并利用 (2.4.9), 我们得到

$$\begin{aligned} r(Q) &= 2m-r\begin{pmatrix} A \\ B \end{pmatrix} \\ &= 2m-r(A)-r(B)+d \\ &\leqslant d. \end{aligned} \tag{2.4.13}$$

类似地可以证明

$$r(Q)\leqslant c. \tag{2.4.14}$$

这就证明了 (2.4.6). 由 (2.4.6) 和 (2.4.10) 可推得 (2.4.5) 的右边, (2.4.4) 和 (2.4.5) 的右边等号成立的条件是显然的. 定理证毕.

推论 2.4.1　(1) $r(A+B)\leqslant r(A\vdots B)$, 等号成立 $\Longleftrightarrow d=r(Q)$;

(2) $r(A+B)\leqslant r\begin{pmatrix} A \\ B \end{pmatrix}$, 等号成立 $\Longleftrightarrow c=r(Q)$,

其中 Q 由 (2.4.3) 所定义.

证明　由 (2.4.7) 和 (2.4.9), 得到

$$r(A+B)=r(A\vdots B)+r(Q)-d.$$

类似地, 由 (2.4.7) 和 (2.4.8) 则可得

$$r(A+B)=r\begin{pmatrix} A \\ B \end{pmatrix}+r(Q)-c,$$

结合 (2.4.6), 所有结论得证.

§2.5 其　　他

本节将讨论两个不等式, 它们各自给出了复方阵秩的一个下界. 这些下界仅仅依赖于矩阵的元素, 因而运算简便, 在不需要计算出矩阵秩的精确值的情形, 有其应用价值.

在下面的讨论中, 我们约定 $\frac{0}{0}=0$.

定理 2.5.1　设 $A=(a_{ij})$ 为 n 阶方阵, 记

$$S_i=\sum_{j=1}^{n}|a_{ij}|,$$

即 S_i 为 A 的第 i 行元素的模之和. 则

$$r(A)\geqslant\sum_{i=1}^{n}\frac{|a_{ii}|}{S_i}.$$

证明　因为用一个不为零的数去乘矩阵的一行并不改变它的秩, 因此不妨设 $a_{ii}\geqslant 0$, 且 $S_i=0$ 或 1. 这样一来, 我们的问题归结为证明

$$r(A)\geqslant \operatorname{tr}(A). \tag{2.5.1}$$

若记 $\lambda_1(A),\cdots,\lambda_n(A)$ 表示 A 的特征值, 则

$$\mathrm{tr}A = \sum_{i=1}^{n} \lambda_i(A) \leqslant \sum_{i=1}^{n} |\lambda_i(A)|. \tag{2.5.2}$$

依 Gerŝgorin 定理 (见定理 4.8.1), 对 A 的任一个特征值 $\lambda(A)$, 存在 $i(1 \leqslant i \leqslant n)$ 使得

$$|\lambda(A) - a_{ii}| \leqslant S_i - a_{ii} \leqslant 1 - a_{ii}.$$

于是

$$|\lambda_i(A)| \leqslant 1, \quad i = 1, \cdots, n.$$

因而 $\sum_{i=1}^{n} |\lambda_i(A)| \leqslant A$ 的非零特征值的个数, 也就是 $r(A)$. 结合 (2.5.2) 知, (2.5.1) 得证. 定理证毕.

定理 2.5.2 设 $A = (a_{ij}) = (a_1, \cdots, a_n)$ 为 n 阶方阵, 这里 $a_1, \cdots, a_n$ 为它的 n 个列向量. 则

$$r(A) \geqslant \sum_{i=1}^{n} \frac{|a_{ii}|^2}{a_i^* a_i}.$$

证明 因为用一个非零数去乘矩阵的一列并不改变它的秩, 我们可假设 $a_i^* a_i = 1$, 于是, 问题归结为证明

$$r(A) \geqslant \sum_{i=1}^{n} |a_{ii}|^2 = \sum_{i=1}^{n} |e_i^* a_i|^2. \tag{2.5.3}$$

其中 $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)'$, 这里 1 为第 i 个元素. 假设 $r(A) = k$, 那么我们可以找到 k 个标准正交化向量 $\varphi_1, \cdots, \varphi_k$, 使得

$$\mathscr{M}(\varphi_1, \cdots, \varphi_k) = \mathscr{M}(A).$$

此时, a_i 有如下表达式

$$a_i = \sum_{j=1}^{k} (\varphi_j^* a_i) \varphi_j,$$

于是

$$e_i^* a_i = \sum_{j=1}^{k} (\varphi_j^* a_i)(e_i^* \varphi_j).$$

应用 Cauchy-Schwarz 不等式及条件 $a_i^* a_i = \sum_{j=1}^{k} |\varphi_j^* a_i|^2 = 1$, 得

$$\begin{aligned}\sum_{i=1}^{n} |e_i^* a_i|^2 &\leqslant \sum_{i=1}^{n} \left(\sum_{j=1}^{k} |\varphi_j^* a_i|^2 \right) \left(\sum_{j=1}^{k} |e_i^* \varphi_j|^2 \right) \\ &= \sum_{j=1}^{k} \sum_{i=1}^{n} |e_i^* \varphi_j|^2 = \sum_{j=1}^{k} |\varphi_j^* \varphi_j|^2 = k = r(A).\end{aligned}$$

(2.5.3) 得证. 定理证毕.

第3章 行 列 式

§3.1 定义及基本性质

定义 3.1.1 n 阶方阵 $A=(a_{ij})$ 的行列式是一个数, 记为 $\det A$, 定义为

$$\det A=\sum_{(i_1,\cdots,i_n)}(-1)^{\tau(i_1,\cdots,i_n)}a_{1i_1}a_{2i_2}\cdots a_{ni_n},$$

其中 $\tau(i_1,\cdots,i_n)$ 表示排列 $(i_1,\cdots,i_n)$ 的逆序数, 和号 $\sum\limits_{(i_1,\cdots,i_n)}$ 表示对 $1,2,\cdots,n$ 的一切排列求和.

行列式有如下性质:

(1) $\det A'=\det A$;

(2) 若 $A_i(i=1,\cdots,k)$ 为 n 阶方阵, c 为任一数, 则

$$\det\prod_{i=1}^{k}A_i=\prod_{i=1}^{k}\det A_i,$$

$$\det(cA)=c^n\det A;$$

(3) 互换行列式的任意两行 (或列), 行列式变号;

(4) 用数乘行列式的某一行 (或列), 等于将行列式乘以该数;

(5) 将行列式的一行 (或列) 乘以一个数, 然后加到另一行 (或列) 的相应元素上, 行列式的值不变;

(6) 若行列式中有一行 (或列) 所有元素全为零, 则行列式等于 0; 若行列式有两行 (或列) 对应元素都相同或成比例, 则行列式为零;

(7) 若行列式的第 i 行 (或列) 的所有元素均可表为两项之和, 则该行列式可表为两个行列式之和, 它们的第 i 行 (或列) 的元素分别是原来行列式的第 i 行 (或列) 的两个加项, 而其余行 (或列) 均与原来的行列式相同.

对 n 阶方阵 $A=(a_{ij})$, 记 M_{ij} 为去掉 A 的第 i 行和第 j 列后所剩的 $n-1$ 阶方阵的行列式. 则

$$A_{ij}=(-1)^{i+j}M_{ij}$$

称为元素 a_{ij} 的代数余子式.

定理 3.1.1 (Laplace 展开)

$$\sum_{j=1}^{n} a_{ij}A_{kj} = \begin{cases} \det A, & 若\ i = k, \\ 0, & 若\ i \neq k, \end{cases}$$

若

$$\sum_{i=1}^{n} a_{ij}A_{ik} = \begin{cases} \det A, & 若\ j = k, \\ 0, & 若\ j \neq k, \end{cases}$$

即 $\det A$ 等于它的任一行 (或列) 的每个元素分别与其对应的代数余子式的乘积之和. 另一方面, A 的任一行 (或列) 的每个元素分别与另一行 (或列) 对应元素的代数余子式的乘积之和等于零.

定理 3.1.2 (Binet-Cauchy) 设 A 和 B 分别为 $m \times n$ 和 $n \times m$ 矩阵, 则

$$\det AB = \begin{cases} 0, & 若\ m > n, \\ \det A \cdot \det B, & 若\ m = n, \\ \displaystyle\sum_{1 \leqslant j_1 < \cdots < j_m \leqslant n} \det A \begin{pmatrix} 1 \cdots m \\ j_1 \cdots j_m \end{pmatrix} \det B \begin{pmatrix} j_1 \cdots j_m \\ 1 \cdots m \end{pmatrix}, & \\ & 若\ m < n, \end{cases}$$

其中 $A\begin{pmatrix} l_1, & \cdots l_t \\ j_1, & \cdots j_t \end{pmatrix}$ 表示由 A 的第 $l_1, \cdots, l_t$ 行和第 $j_1, \cdots, j_t$ 列交叉处元素组成的子阵.

下面的两个定理在后面的讨论中多次用到.

定理 3.1.3 将方阵 A 分块为

$$A = \begin{pmatrix} A_{11}, & A_{12} \\ A_{21}, & A_{22} \end{pmatrix},$$

(1) 若 A_{11} 可逆, 则 $\det A = \det A_{11} \det A_{22.1}$;

(2) 若 A_{22} 可逆, 则 $\det A = \det A_{22} \det A_{11.2}$,

其中 $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}, A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

证明 在下式两边取行列式即得 (1).

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}.$$

类似可证明 (2).

定理 3.1.4　设 A 和 B 分别为 $n\times m$ 和 $m\times n$ 矩阵, 则

$$\det(I_n - AB) = \det(I_m - BA).$$

此即推论 1.2.4.

§3.2　半正定阵之和的行列式

设 A 和 B 皆为半正定的 Hermite 阵, 本节将讨论有关 $\det(A+B)$ 的各种不等式.

定理 3.2.1　设 Hermite 阵 $A\geqslant 0, B\geqslant 0$. 则

$$\det(A+B)\geqslant \det A + \det B, \tag{3.2.1}$$

等号成立 $\Longleftrightarrow$ $A=0$ 或 $B=0$ 或 $\det(A+B)=0$.

证明　根据定理 1.4.1(2), 存在酉阵 P, 使 $PBP^*=\Lambda$, 其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, $\lambda_1\geqslant\cdots\geqslant\lambda_n\geqslant 0$ 为 B 的特征值. 因为 $\det(A+B)=\det(PAP^*+\Lambda)$, $\det A=\det PAP^*$, $\det B=\det\Lambda$, 且 $PAP^*\geqslant 0$, 故不失普遍性, 我们可假设 $B=\Lambda$.

将 $\det(A+\Lambda)$ 展开, 得

$$\det(A+\Lambda)=\det A+\sum_{i=1}^{n}\lambda_i d_i+\sum_{1\leqslant i<j\leqslant n}\lambda_i\lambda_j d_{ij}+\cdots+\prod_{i=1}^{n}\lambda_i, \tag{3.2.2}$$

其中 $d_{i_1 i_k}$ 表示从 A 中剔除第 $i_1,\cdots,i_k$ 行和列之后剩下的方阵的行列式, 因为 $A\geqslant 0$, 故所有的 $d_{i_1\cdots i_k}\geqslant 0$(见定理 1.4.2(5)). 又因 $\det B=\prod_{i=1}^{n}\lambda_i$, 从 (3.2.2) 我们得到

$$\det(A+\Lambda)\geqslant \det A+\det\Lambda. \tag{3.2.3}$$

于是 (3.2.1) 得证.

现在证明等号成立的充要条件. 充分性是显然的, 以下证必要性, 分两种情况来考虑.

(1) 若 $\det B\neq 0$, 此时每个 $\lambda_i\neq 0$, 从 (3.2.2) 知, 在 (3.2.3) 中等号成立, 必有 $d_{i_1\cdots i_k}=0$, 对一切 $i_1,\cdots,i_k$. 特别当 $k=n-1$ 时, $d_{i_1\cdots i_{n-1}}$ 就是 A 的对角元. 这就证明了 A 的所有对角元为零. 但 $A\geqslant 0$, 故 $A=0$.

(2) 若 $\det B=0$, 且 $B\neq 0$, 此时至少有一个 $\lambda_i\neq 0$. 因所有 $d_{i_1\cdots i_k}\geqslant 0$, 但从 (3.2.2) 知, 至少有一个 $d_{i_1\cdots i_k}=0$. 因 $A\geqslant 0$, 根据定理 1.4.3(5), A 不可能是正定的, 于是 $\det A=0$. 因为我们假设了 $\det(A+B)=\det A+\det B$, 所以, $\det(A+B)=0$. 证毕.

注 1 这个定理有很多证明方法, 此处证法取自华罗庚 (1955).

推论 3.2.1 设 Hermite 阵 A 和 B 皆为半正定阵, 且 $A \geqslant B$, 即 $A-B \geqslant 0$. 则

$$\det A \geqslant \det B.$$

证明 记 $C=A-B \geqslant 0$. 因此 $\det C \geqslant 0$. 应用定理得

$$\det A=\det(B+C) \geqslant \det B+\det C \geqslant \det B.$$

证毕.

为了进一步讨论的需要, 我们现在引进 Schur 补矩阵的概念.

定义 3.2.1 将方阵 A 分块为

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{3.2.4}$$

当 A_{11} 可逆时, 称方阵

$$A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12} \tag{3.2.5}$$

为 A_{11} 在 A 中的 Schur 补矩阵, 简称Schur 补, 也常记为

$$A_{22.1}=(A/A_{11}). \tag{3.2.6}$$

引理 3.2.1 设 A 由 (3.2.4) 定义. 则

(1) 当 A 为 Hermite 阵时, $A>0 \Longleftrightarrow A_{11}>0, (A/A_{11})>0$;

(2) 若 A 为 Hermite 阵且 $A_{11}>0$. 则 $A \geqslant 0 \Longleftrightarrow (A/A_{11}) \geqslant 0$;

(3) $\det A=\det A_{11}\det(A/A_{11})$.

证明 (1) 因 $A>0$ 时 $A_{11}>0$, 故我们可在假设 $A_{11}>0$ 下, 证明本命题. 因为 A_{11} 可逆, 故

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}=\begin{pmatrix} A_{11} & 0 \\ 0 & (A/A_{11}) \end{pmatrix}, \tag{3.2.7}$$

应用定理 1.4.3(4) 知, $A>0 \Longleftrightarrow \operatorname{diag}(A_{11},(A/A_{11}))>0 \Longleftrightarrow A_{11}>0$ 和 (A/A_{11}) >0.

(2) 的证明与 (1) 类似.

(3) 为定理 3.1.3 (1) 的另一种叙述.

引理证毕.

由引理中的 (3), 立即得到

$$\det(A/A_{11})=\frac{\det A}{\det A_{11}}. \tag{3.2.8}$$

类似地, 若 A_{22} 可逆, 则可定义 A_{22} 在 A 中的 Schur 补

$$(A/A_{22}) = A_{11} - A_{12}A_{22}^{-1}A_{21},$$

于是, 我们有

$$\det(A/A_{22}) = \frac{\det A}{\det A_{22}}.$$

引理 3.2.2　设

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

皆为 $m+n$ 阶 Hermite 阵, 其中 A_{11} 和 B_{11} 为 m 阶方阵, 若 $A \geqslant 0$, $B \geqslant 0$. $A_{11} > 0, B_{11} > 0$. 则

$$((A+B)/(A_{11}+B_{11})) \geqslant (A/A_{11}) + (B/B_{11}). \tag{3.2.9}$$

证明　因为

$$A+B = \begin{pmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{pmatrix} \geqslant 0,$$
$$A_{11}+B_{11} > 0.$$

由上一引理 (2) 知, $A_{11}+B_{11}$ 的 Schur 补

$$\begin{aligned} &((A+B)/(A_{11}+B_{11})) \\ &= (A_{22}+B_{22}) - (A_{21}+B_{21})(A_{11}+B_{11})^{-1}(A_{12}+B_{12}) \geqslant 0. \end{aligned}$$

于是

$$\begin{aligned} &((A+B)/(A_{11}+B_{11})) - (A/A_{11}) - (B/B_{11}) \\ &= A_{21}A_{11}^{-1}A_{12} + B_{21}B_{11}^{-1}B_{12} - (A_{21}+B_{21})(A_{11}+B_{11})^{-1}(A_{12}+B_{12}) \\ &= (A_{21} \vdots B_{21}) \begin{pmatrix} A_{11}^{-1} - (A_{11}+B_{11})^{-1} & -(A_{11}+B_{11})^{-1} \\ -(A_{11}+B_{11})^{-1} & B_{11}^{-1} - (A_{11}+B_{11})^{-1} \end{pmatrix} \begin{pmatrix} A_{12} \\ B_{12} \end{pmatrix} \\ &= (A_{21} \vdots B_{21}) \begin{pmatrix} I \\ -B_{11}^{-1}A_{11} \end{pmatrix} (A_{11} + A_{11}B_{11}^{-1}A_{11})^{-1} (I \vdots -A_{11}B_{11}^{-1}) \begin{pmatrix} A_{12} \\ B_{12} \end{pmatrix} \\ &\geqslant 0. \end{aligned}$$

在最后一步, 应用了恒等式 $(A_{11} + A_{11}B_{11}^{-1}A_{11})^{-1} = A_{11}^{-1} - (A_{11}+B_{11})^{-1}$. 引理证毕.

定理 3.2.2 设 A 和 B 为同阶 Hermite 阵,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \geqslant 0, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \geqslant 0,$$

且 $A_{11} > 0$, $B_{11} > 0$, 这里 A_{11} 与 B_{11} 具有相同阶. 则

$$\frac{\det(A+B)}{\det(A_{11}+B_{11})} \geqslant \frac{\det A}{\det A_{11}} + \frac{\det B}{\det B_{11}}. \tag{3.2.10}$$

证明 应用 (3.2.8), 引理 3.2.2, 推论 3.2.1 及定理 3.2.1 得

$$\begin{aligned}\frac{\det(A+B)}{\det(A_{11}+B_{11})} &= \det((A+B)/(A_{11}+B_{11})) \\ &\geqslant \det((A/A_{11}) + (B/B_{11})) \\ &\geqslant \det(A/A_{11}) + \det(B/B_{11}) \\ &= \frac{\det A}{\det A_{11}} + \frac{\det B}{\det B_{11}}.\end{aligned} \tag{3.2.11}$$

证毕.

推论 3.2.2 (Bergstrom 不等式) 设 A 和 B 为同阶正定 Hermite 阵. 用 $A_{(i)}$ 表示剔除 A 的第 i 行和第 i 列后剩下的主子阵. 则

$$\frac{\det(A+B)}{\det(A_{(i)}+B_{(i)})} \geqslant \frac{\det A}{\det A_{(i)}} + \frac{\det B}{\det B_{(i)}}, \quad i = 1, \cdots, n.$$

注 2 根据定理 1.4.3(4), 当 A, B 为正定 Hermite 阵时, 它们的任一个主子阵也是正定的, 因此推论中 $A_{(i)}$ 和 $B_{(i)}$ 换成 A 和 B 的任意 k 阶主子阵 $A(i_1, \cdots, i_k)$ 和 $B(i_1, \cdots, i_k)$, 不等式仍然成立.

为符号简单计, 以下用 A_i 表示方阵 A 的左上角 i 阶子方阵.

定理 3.2.3 设 n 阶 Hermite 阵 A 和 B 皆为半正定的, 且 $A_i > 0$, $B_i > 0, i = 1, \cdots, n-1$. 则

$$\begin{aligned}\det(A+B) \geqslant{}& \det A \cdot \left(1 + \sum_{i=1}^{n-1} \frac{\det B_i}{\det A_i}\right) + \det B \cdot \left(1 + \sum_{i=1}^{n-1} \frac{\det A_i}{\det B_i}\right) \\ &+ (2^n - 2n)(\det A \det B)^{1/2}.\end{aligned} \tag{3.2.12}$$

证明 对阶数 n 用归纳法. 当 $n = 2$ 时, 由引理 3.2.1(3) 得

$$\det(A+B) = \det(A_1 + B_1) \cdot \det((A+B)/(A_1+B_1)).$$

显然

$$\det(A_1+B_1)\geqslant \det A_1+\det B_1.$$

再应用定理 3.2.2,

$$\det((A+B)/(A_1+B_1))\geqslant \frac{\det A}{\det A_1}+\frac{\det B}{\det B_1},$$

综合上述三式, 命题 $n=2$ 得证.

现在假设 (3.2.12) 对 $n\times n$ 半正定 Hermite 阵 A 和 B 成立, 又设 $\bar{A}$ 和 $\bar{B}$ 为 $(n+1)\times(n+1)$ 半正定 Hermite 阵, 且以 A 和 B 分别为它们的左上角 $n\times n$ 子阵, 记 $A_n=A, B_n=B$. 则由引理 3.2.1(3) 得

$$\begin{aligned}\det(\bar{A}+\bar{B})&=\det(A_n+B_n)\cdot\det((\bar{A}+\bar{B})/(A_n+B_n))\\&=\det(A+B)\cdot\det((\bar{A}+\bar{B})/(A+B)).\end{aligned}\tag{3.2.13}$$

而由定理 3.2.2, 有

$$\det((\bar{A}+\bar{B})/(A+B))\geqslant\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}.$$

对 (3.2.13) 第一因子应用归纳假设, 结合上式, 我们得到

$$\begin{aligned}\det(\bar{A}+\bar{B})\geqslant&\left[\det A\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}\right)+\det B\left(1+\sum_{i=1}^{n-1}\frac{\det A_i}{\det B_i}\right)\right.\\&\left.+(2^n-2n)(\det A\det B)^{1/2}\right]\left(\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}\right)\\=&\det\bar{A}\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}\right)+\frac{\det B}{\det A}\det\bar{A}\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}\right)\\&+\frac{\det A}{\det B}\det\bar{B}\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}\right)\\&+\det\bar{B}\left(1+\sum_{i=1}^{n-1}\frac{\det A_i}{\det B_i}\right)+(2^n-2n)(\det A\det B)^{1/2}\\&\cdot\left(\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}\right)\\=&\det\bar{A}\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}\right)+\frac{\det B}{\det A}\det\bar{A}+\det\bar{B}\\&\cdot\left(1+\sum_{i=1}^{n-1}\frac{\det A_i}{\det B_i}\right)+\frac{\det A}{\det B}\det\bar{B}\\&+(2^n-2n)(\det A\det B)^{1/2}\left(\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}\right)\end{aligned}$$

$$+\sum_{i=1}^{n-1}\left(\frac{\det A_i}{\det B_i}\frac{\det B}{\det A}\det\bar{A}+\frac{\det B_i}{\det A_i}\frac{\det A}{\det B}\det\bar{B}\right)$$

$$=\det\bar{A}\left(1+\sum_{i=1}^{n}\frac{\det B_i}{\det A_i}\right)+\det\bar{B}\left(1+\sum_{i=1}^{n}\frac{\det A_i}{\det B_i}\right)$$

$$+(2^n-2n)(\det A\det B)^{1/2}\left(\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}\right)$$

$$+\sum_{i=1}^{n-1}\left(\frac{\det A_i}{\det B_i}\frac{\det B}{\det A}\det\bar{A}+\frac{\det B_i}{\det A_i}\frac{\det A}{\det B}\det\bar{B}\right). \tag{3.2.14}$$

记上式的第三项为 Δ. 利用如下事实：假设 $a>0, b>0$, 对一切 $0<x<\infty$, 有

$$ax+\frac{b}{x}\geqslant 2\sqrt{ab}, \tag{3.2.15}$$

易证

$$\Delta\geqslant 2(n-1)(\det\bar{A}\det\bar{B})^{1/2}, \tag{3.2.16}$$

再利用几何平均与算术平均不等式, 有

$$\frac{\det\bar{A}}{\det A}+\frac{\det\bar{B}}{\det B}\geqslant 2\left(\frac{\det\bar{A}}{\det A}\frac{\det\bar{B}}{\det B}\right)^{1/2}. \tag{3.2.17}$$

将 (3.2.16) 和 (3.2.17) 代入 (3.2.14), 得到

$$\det(\bar{A}+\bar{B})\geqslant\det\bar{A}\left(1+\sum_{i=1}^{n}\frac{\det B_i}{\det A_i}\right)+\det\bar{B}\left(1+\sum_{i=1}^{n}\frac{\det A_i}{\det B_i}\right)$$
$$+[2^{n+1}-2(n+1)](\det\bar{A}\det\bar{B})^{1/2}.$$

这就完成了归纳法证明. 定理证毕.

推论 3.2.3 设 A 和 B 为两个 n 阶 Hermite 阵, $A>0, B>0$, 则

$$\det(A+B)\geqslant\det A+\det B+(2^n-2)(\det A\det B)^{1/2}. \tag{3.2.18}$$

证明 因为 $A>0, B>0$. 所以它们的所有主子阵都是正定的. 由定理 3.2.3, 得

$$\det(A+B)\geqslant\det A+\det B+\sum_{i=1}^{n-1}\left(\det A\frac{\det B_i}{\det A_i}+\det B\frac{\det A_i}{\det B_i}\right)$$
$$+(2^n-2n)(\det A\det B)^{1/2}.$$

记上式第三项整个和式为 Δ, 应用 (3.2.15), 我们有 $\Delta\geqslant 2(n-1)(\det A\det B)^{1/2}$. 代入上式, 命题得证.

比较不等式 (3.2.1) 和 (3.2.18), 我们看到, 对正定的 Hermite 阵, (3.2.18) 是对 (3.2.1) 的一个改进.

推论 3.2.4　设 A 和 B 为 n 阶 Hermite 阵且 $A>0, B>0, A>B$(即 $A-B>0$). 则

$$\det(A+B)>\det A+(2^n-2)\det B. \tag{3.2.19}$$

证明　因为 $A>0, B>0, A>B$, 类似于推论 3.2.1 可以证明 $\det A>\det B$, 于是 $(\det A)^{1/2}>(\det B)^{1/2}$. 再结合 (3.2.18) 式, 命题得证.

推论 3.2.5　在推论 3.2.4 的条件下, 有

$$\det(A+B)>\det A+n\det B.$$

证明　沿用定理 3.2.3 的记号. 由 $A-B>0$, 知它的子方阵 $A_i-B_i>0$, $i=1,\cdots,n$, 因而 $\det A_i>\det B_i$. 应用 (3.2.12) 得

$$\begin{aligned}\det(A+B)&\geqslant\det A\left(1+\sum_{i=1}^{n-1}\frac{\det B_i}{\det A_i}+n\det B\right)\\&>\det A+n\det B.\end{aligned}$$

最后, 在结束这一节的时候, 我们证明著名的 Minkowski 不等式.

定理 3.2.4 (Minkowski)　设 A 和 B 为 n 阶正定 Hermite 阵. 则

$$(\det(A+B))^{1/n}\geqslant(\det A)^{1/n}+(\det B)^{1/n}, \tag{3.2.20}$$

且等号成立 $\Longleftrightarrow B=aA$, 这里 $a>0$.

证明　不失一般性, 可以假设 $A=I$, 于是问题归结为证明

$$(\det(I_n+B))^{1/n}\geqslant1+(\det B)^{1/n}. \tag{3.2.21}$$

设 $\lambda_1\cdots\lambda_n(\lambda_1\geqslant\cdots\geqslant\lambda_n>0)$ 为 B 的特征值, 则上面的不等式等价于

$$\prod_{i=1}^{n}(1+\lambda_i)\geqslant(1+\sqrt[n]{\lambda_1\cdots\lambda_n})^n, \tag{3.2.22}$$

将两边直接乘开来, 并用算术平均与几何平均不等式, 即证 (3.2.22).

在 (3.2.22) 中, 等号成立当且仅当 $\lambda_1=\cdots=\lambda_n=a$, 这里 a 为任一正数, 这就是说, 在 (3.2.21) 中等号成立的充要条件为 $B=aI_n$. 对应到 (3.2.20), 即 $B=aA$. 定理证毕.

注 3　当 A 和 B 为实矩阵时, Minkowski 不等式能够推广到非对称阵的情形, 关于这方面的结果读者可参阅 Haynesworth(1957, 1960).

令 $A(i_1,\cdots,i_k)$ 表示由 A 的第 $i_1,\cdots,i_k$ 行和列交叉处元素组成的主子阵. 记 $\alpha=(i_1,\cdots,i_k)$, $A(\alpha)=A(i_1,\cdots,i_k)$. 若 $\alpha=(l,l+1,\cdots,m)$, 则记 $A(\alpha)=A_{l,m}$. 若 $\alpha,\beta\in\{1,\cdots,n\}$, $\alpha\cap\beta=\varnothing$, 则记 $|A(\alpha\cup\beta)|=1$.

以下定理的证明主要应用了 Jensen 不等式的逆矩阵形式, 由于涉及知识比较广, 证明从略, 感兴趣的读者可参阅 Malamud(2001).

定理 3.2.5(Hadamard-Fischer 不等式) 设 A 为 n 阶正定 Hermite 阵, α, $\beta\subset\{1,\cdots,n\}$, 则

$$|A(\alpha\cup\beta)|\leqslant\frac{|A(\alpha)|\cdot|A(\beta)|}{|A(\alpha\cap\beta)|}. \tag{3.2.23}$$

定理 3.2.6 设 A 为 n 阶正定 Hermite 阵, 若 $\alpha\cup\beta=\{1,\cdots,n\}$, $\alpha\setminus\beta$ 含有 k 元素, $1\leqslant k\leqslant n$, 则

$$|A|\leqslant\frac{|A(\alpha)|\cdot|A(\beta)|}{|A(\alpha\cap\beta)|}\leqslant\frac{(b_k+a_k)^2}{4b_ka_k}|A|, \tag{3.2.24}$$

特别地

$$|A|\leqslant|A_{1,k}|\cdot|A_{k+1,n}|\leqslant\frac{(b_k+a_k)^2}{4b_ka_k}, \tag{3.2.25}$$

这里 $b_k=\lambda_1\cdots\lambda_k$, $a_k=\lambda_{n-k+1}\cdots\lambda_n$, 其中 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 为 A 的特征值.

定理 3.2.7 采用定理 3.2.6 的记号, 我们有

$$0\leqslant\frac{|A(\alpha)|}{|A|}-\frac{|A(\alpha\cap\beta)|}{|A(\beta)|}\leqslant\frac{(\sqrt{b_k}-\sqrt{a_k})^2}{b_ka_k}. \tag{3.2.26}$$

定理 3.2.8 设 A 为 n 阶正定 Hermite 阵, 其特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$, $1\leqslant k\leqslant n$, 则

$$\lambda_{n-k+1}\cdots\lambda_n\leqslant\frac{|A|}{|A_{k+1,n}|}\leqslant\prod_{i=1}^{k}\frac{|A(i,k+1,\cdots,n)|}{|A_{k+1,n}|}\leqslant\Big(\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{k}\Big)^k. \tag{3.2.27}$$

定理 3.2.9 设 A 和 B 为 n 阶正定 Hermite 阵, $A\geqslant B>0$, 则

$$\frac{|A|}{|A_{k+1,n}||B_{1,k}|}\leqslant\frac{|A+B|}{|A_{1,k}+B_{1,k}||A_{k+1,n}+Bk+1,n|}\leqslant\frac{|A|}{|A_{1,k}||B_{k+1,n}|}. \tag{3.2.28}$$

§3.3 Hadamard 不等式

对于正定阵, Hadamard 不等式是一个很基本的不等式, 迄今为止已积累了上百种不同的证明方法 (见 Bellman(1970).p.140). 本节只给出一种证法, 在 §8.3 我们

还将叙述另一种证法. 这个不等式之所以重要是因为许多其他不等式都可以从它导出.

定理 3.3.1 (Hadamard)　设 n 阶 Hermite 阵 $A=(a_{ij})\geqslant 0$. 则

$$\det A\leqslant \prod_{i=1}^{n} a_{ii}, \tag{3.3.1}$$

且等号成立 $\Longleftrightarrow$ A 为对角阵.

证明　若 $\det A=0$, 不等式 (3.3.1) 自然成立. 若 $\det A\neq 0$, 则 $a_{ii}>0, i=1,\cdots,n$. 记 $D=\mathrm{diag}(d_i,\cdots,d_n)$, $d_i=a_{ii}^{-1/2}, i=1,\cdots,n, B=DAD$. 依定理 1.4.3(4) 知 $B>0$, 易见 (3.3.1) 等价于

$$\det B\leqslant 1. \tag{3.3.2}$$

注意, B 的对角元全为 1.

现在我们证明 (3.3.2). 记 $\lambda_i(B)(i=1,\cdots,n)$ 为 B 的特征值, 应用算术平均与几何平均不等式, 得

$$\det B\prod_{i=1}^{n}\lambda_i(B)\leqslant\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i(B)\right)^n=\left(\frac{1}{n}\mathrm{tr}B\right)^n=1, \tag{3.3.3}$$

且等号成立 $\Longleftrightarrow$ 所有 $\lambda_i(B)=1$, 因为 B 也是 Hermite 阵, 所以 (3.3.3) 等号成立 $\Longleftrightarrow$ $B=I$, 这等价于 A 为对角阵. 证毕.

下一个行列式的不等式是对一般方阵而言的, 也常常称为 Hadamard 不等式.

推论 3.3.1　对任意的 $n\times m$ 矩阵 $A=(a_{ij})$,

$$\det AA^*\leqslant\prod_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2, \tag{3.3.3}$$

且等号成立 $\Longleftrightarrow$ A 的行向量互相正交.

证明　对半正定 Hermite 阵 $B=AA^*$, 应用定理 3.3.1, 得

$$\det AA^*=\det B\leqslant\prod_{i=1}^{n}b_{ii}=\prod_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2, \tag{3.3.4}$$

于是 (3.3.3) 得证. 注意到, (3.3.4) 中等号成立, 当且仅当 B 为对角阵, 这等价于 A 的行向量互相正交, 推论证毕.

注　当 A 为 $n\times n$ 方阵时, (3.3.3) 可改写为

$$|\det A|\leqslant\prod_{i=1}^{n}\left(\sum_{i=1}^{m}|a_{ij}|^2\right)^{1/2}.$$

当 A 为实方阵时, 这个不等式的几何意义如下, 记 $a'_{(1)},\cdots,a'_{(n)}$ 为 A 的行向量, 即 $A'=(a_{(1)},\cdots,a_{(n)})$, 则上式变形为

$$|\det A|\leqslant\prod_{i=1}^{n}\|a_{(i)}\|, \tag{3.3.5}$$

$|\det A|$ 表示以 $a_{(1)},\cdots,a_{(n)}$ 为棱的超平行六面体的体积, 而不等式右边是以 $a_{(1)},\cdots,a_{(n)}$ 为棱的超长方体的体积. (3.3.5) 是如下简单几何事实的一个代数描述：一个超平行六面体的体积不超过各条棱具有相同长度的超长方体的体积.

推论 3.3.2 设 $A=(a_{ij})$ 为任意 n 阶方阵, 记 $M=\max\{|a_{ij}|\}$. 则

$$|\det A|\leqslant M^n n^{n/2}.$$

这里推论 3.3.1 的直接结果.

应用 Hadamard 不等式我们可以证明下面的所谓 $(\det A)^{1/n}$ 表示定理.

定理 3.3.2 设 A 为 n 阶正定 Hermite 阵. 则

$$(\det A)^{1/n}=\min\frac{1}{n}\mathrm{tr}(AX), \tag{3.3.6}$$

其中 X 为正定 Hermite 阵, 且 $\det X=1$.

证明 根据定理 1.4.1, 存在酉阵 U, 使得 $A=U\Lambda U^*$, 这里 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, $\lambda_i>0, i=1,\cdots,n$ 于是

$$\mathrm{tr}(AX)=\mathrm{tr}(\Lambda U^*XU)=\mathrm{tr}(\Lambda D),$$

其中 $D=U^*XU=(d_{ij})>0$, 故 $d_{ii}>0, i=1,\cdots,n$. 应用算术平均与几何平均不等式及 Hadamard 不等式, 得

$$\begin{aligned}\frac{1}{n}\mathrm{tr}(AX)&=\frac{1}{n}\sum_{i=1}^{n}\lambda_i d_{ii}\geqslant\left(\prod_{i=1}^{n}\lambda_i d_{ii}\right)^{1/n}=(\det A)^{1/n}\left(\prod_{i=1}^{n}d_{ii}\right)^{1/n}\\&\geqslant(\det A)^{1/n}(\det D)^{1/n}=(\det A)^{1/n},\end{aligned}$$

最后一个等号是因为 $\det D=\det X=1$. 从证明过程知, 当

$$X=U\mathrm{diag}(d_{11},\cdots,d_{nn})U^*, \tag{3.3.7}$$

且 $\lambda_i \boldsymbol{d}_{11}=\cdots=\lambda_n d_{nn}, \det X=1$ 时, (3.3.6) 的右边达到最小值. 这要求 (3.3.7) 中

$$d_{ii}=\frac{(\det A)^{1/n}}{\lambda_i},\qquad i=1,\cdots,n.$$

定理证毕.

为了研究 Hadamard 不等式的精确度, Dixon(1984) 研究了 Hadamard 商

$$h(A)=\frac{|\det A|}{\prod_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}.$$

他证明了, 若 A 为实随机阵, 它的分布满足一些简单的对称性质, 那么对随机变量 $\ln h(A)$, 有

$$E\ln h(A)=-\frac{1}{2}n-\frac{1}{4}\ln n+O(1),$$

$$\operatorname{var}(\ln h(A))=\frac{1}{2}\ln n+O(1),$$

这里 $E(\cdot)$ 和 $\operatorname{var}(\cdot)$ 表均值和方差.

§3.4　Fischer 不等式

下面的 Fischer 不等式可以看作 Hadamard 不等式的推广, 在这里不等式 (3.3.1) 右边的对角元用主子式所代替.

定理 3.4.1 (Fischer 不等式)　设 Hermite 阵

$$A=\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{12} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{pmatrix}\geqslant 0,$$

这里 $A_{ii}(i=1,\cdots,k)$ 皆为方阵, 则

$$\det A\leqslant\prod_{i=1}^{k}\det A_{ii}, \tag{3.4.1}$$

且等号成立 $\Longleftrightarrow$ A 为准对角阵, 即 $A_{ij}=0, i\neq j$.

证明　易见, 我们只需对 $A>0$, $k=2$ 证明 (3.4.1). 因 $A>0$ 时, $A_{11}>0$, 且

$$A_{22}\geqslant A_{22}-A_{21}A_{11}^{-1}A_{12}=(A/A_{11}).$$

于是由推论 3.2.1.

$$\det A_{22}\geqslant\det(A/A_{11}),$$

结合引理 3.2.1(3), 有

$$\det A=\det A_{11}\det(A/A_{11})\leqslant\det A_{11}\det A_{22}.$$

定理证毕.

注 若将 $A=(a_{ii})\geqslant 0$ 看作本定理 $A_{ij}=a_{ij}$ 的分块, Fischer 不等式就变为 Hadamard 不等式.

应用本定理容易证明下面的事实.

定理 3.4.2 设 A 为 $n\times n$ 矩阵, 将其分块为 $A=(A_1\vdots A_2)$. 则

$$|\det A|^2\leqslant \det A_1^*A_1\cdot\det A_2^*A_2,$$

且等号成立 $\Longleftrightarrow A_1^*A_2=0$.

证明 对

$$A^*A=\begin{pmatrix} A_1^*A_1 & A_1^*A_2\\ A_2^*A_1 & A_2^*A_2\end{pmatrix}$$

应用定理 3.4.1, 即得欲证.

§3.5 Szasz 不等式

设 A 为 $n\times n$ 方阵, $A(i_1,\cdots,i_k)$ 表示它的第 $i_1,\cdots,i_k$ 行和列交叉处元素组成的主子阵, $\det A(i_1,\cdots,i_k)$ 就是对应的主子式. 用 $P_k(A)$ 表示 A 的所有 k 阶主子式的乘积, 即

$$P_k(A)=\prod_{1\leqslant i_1<\cdots<i_k\leqslant n}\det A(i_1,\cdots,i_k).$$

这样的乘积共有 $\begin{pmatrix} n\\ k\end{pmatrix}$ 个. 特别, $P_n(A)=\det A, P_1(A)=\prod\limits_{i=1}^{n}a_{ii}$.

定理 3.5.1 (Szasz) 设 n 阶 Hermite 阵 $A>0$. 则

$$[P_{k+1}(A)]^{\binom{n-1}{k}^{-1}}\leqslant [p_k(A)]^{\binom{n-1}{k-1}^{-1}},\quad k=1,\cdots,n-1. \tag{3.5.1}$$

证明 因为 A^{-1} 的对角元是 A 的 $(n-1)\times(n-1)$ 阶主子式与 $\det A$ 之商, 应用 Hadamard 不等式, 有

$$\frac{1}{\det A}=\det A^{-1}\leqslant\frac{P_{n-1}(A)}{(\det A)^n},$$

于是

$$[P_n(A)]^{n-1}=(\det A)^{n-1}\leqslant P_{n-1}(A), \tag{3.5.2}$$

即

$$P_n(A)\leqslant (P_{n-1}(A))^{1/(n-1)}. \tag{3.5.3}$$

这就对 $k=n-1$ 证明了 (3.5.1).

对 $k=n-2$, 我们用归纳式来证明, 此时把每个 $n-1$ 阶主子阵看作前面讨论的初始矩阵 A. 应用 (3.5.2), 并注意到 A 的每个 $(n-2)\times(n-2)$ 的主子阵可以看作 A 的两个不同的 $(n-1)\times(n-1)$ 阶主子阵的主子阵, 于是, 我们得到

$$[P_{n-1}(A)]^{n^{-2}} \leqslant (P_{n-2}(A))^2,$$

即

$$(P_{n-1}(A))^{1/(n-1)} \leqslant (P_{n-2}(A))^{2/(n-1)(n-2)},$$

这就对 $k=n-2$ 证明了 (3.5.1). 定理证毕.

注　(3.5.1) 可改写为

$$\begin{aligned}\det A = P_n(A) &\leqslant (P_{n-1}(A))^{\binom{n-1}{n-2}^{-1}} \leqslant \cdots \leqslant (P_3(A))^{\binom{n-1}{2}^{-1}}\\ &\leqslant (P_2(A))^{\binom{n-1}{1}^{-1}} \leqslant P_1(A).\end{aligned}$$

下面的推论是 (3.5.2) 的一个变形, 它给出了 $\det A$ 的一个上界.

推论 3.5.1　设 n 阶 Hermite 阵 $A>0$. 则

$$\det A \leqslant \left(\prod_{i=1}^{n} \det A_{(i)}\right)^{1/(n-1)},$$

这里 $A_{(i)}$ 表示剔除 A 的第 i 行和第 i 列后剩下的子阵, 即 $A(1,\cdots,i-1,i+1,\cdots,n)$.

§3.6　Oppenhein 不等式

我们先证明一个引理.

引理 3.6.1　设 n 阶 Hermite 阵 $A\geqslant 0$. 则 $A-\alpha E_{11}\geqslant 0$, 其中

$$\alpha = \begin{cases} \dfrac{\det A}{M_{11}}, & 若\ M_{11}\neq 0,\\ 0, & 若\ M_{11}=0,\end{cases}$$

其中 M_{11} 表示将 A 的第 1 行和第 1 列剔除后剩下的 $n-1$ 阶方阵的行列式, E_{11} 表示 (1,1) 元为 1 其余元素皆为零的 n 阶方阵.

证明　(1) 先考虑 $A>0$ 的情况. 记 B 为将 A 的第 $1,\cdots,n$ 行和第 $1,\cdots,n$ 列变成第 $n,\cdots,1$ 行和第 $n,\cdots,1$ 列之后的方阵. 因为将 A 做这样的变化相当于用一个可逆实矩阵 P 左乘 A 和同时用 P' 右乘 A, 故 $B=PAP'$, 且 $\det B=\det A$. 根据定理 1.4.3(4) 知 $B>0$. 因为同样变化之后, $A-cE_{11}$ 变成 $B-cE_{nn}$, 这里 E_{nn} 表示 (n,n) 元为 1 其余元素皆为零的方阵, 所以问题归结为证明 $B-cE_{nn}\geqslant 0$. 注意

到 $B-cE_{nn}$ 和 B 的阶数不超过 $n-1$ 的顺序主子式都相同, 所以 $B-cE_{nn}$ 的阶数不超过 $n-1$ 的顺序主子式都是正数, 应用行列式性质 (7) 可以推得

$$\begin{aligned}\det(B-cE_{nn})&=\det B-cM_{11}\\&=\det A-cM_{11}>0,\ \text{当}\ c<\alpha\ \text{时}.\end{aligned}$$

故 $A-cE_{11}>0$, 当 $c<\alpha$ 时. 这就证明了 $A-\alpha E_{11}\geqslant 0$.

(2) 若 $A\geqslant 0$, 则对 $\varepsilon>0, A+\varepsilon I>0$. 用 $A+\varepsilon I$ 代替 A, 重复上面的讨论, 可以证得 $(A+\varepsilon I)-\alpha(\varepsilon)E_{11}\geqslant 0$. 令 $\varepsilon\to 0$, 即得所证. 证毕.

利用 Hadamard 乘积 (见 §1.10), Hadamard 不等式 (3.3.1) 可以表为 $\det A \leqslant \det(A\circ I)$, 即

$$\det A\cdot\prod_{i=1}^{n}1\leqslant\det(A\circ I). \tag{3.6.1}$$

下面的 Oppenhein 不等式是 (3.6.1) 的一个推广.

定理 3.6.1 (Oppenhein) 设 n 阶 Hermite 阵 $A>0, B>0$. 则

$$\det A\cdot\prod_{i=1}^{n}b_{ii}\leqslant\det(A\circ B). \tag{3.6.2}$$

证明 对阶数 n 施归纳法. $n=1$ 命题显然成立. 现设 $n\geqslant 2$, 且 (3.6.2) 对一切小于或等于 $n-1$ 阶的正定 Hermite 阵都成立. 记 A_{11} 和 B_{11} 为将 A 和 B 的第 1 行和第 1 列剔除后剩下的 $n-1$ 阶主子阵, 依归纳假设, 有

$$\det A_{11}\prod_{i=2}^{n}b_{ii}\leqslant\det(A_{11}\circ B_{11}). \tag{3.6.3}$$

设 α 同引理, 因 $A-\alpha E_{11}\geqslant 0$, 应用定理 1.10.1(1), 得

$$(A-\alpha E_{11})\circ B\geqslant 0.$$

因

$$\begin{aligned}(A-\alpha E_{11})\circ B&=A\circ B-\alpha E_{11}\circ B\\&=\left\{\begin{array}{c:c}a_{11}b_{11}-\alpha b_{11}&a_{12}b_{12}\cdots a_{1n}b_{1n}\\\hdashline a_{21}b_{21}+0&\\\vdots&A_{11}\circ B_{11}\\a_{n1}b_{n1}+0&\end{array}\right\},\end{aligned}$$

应用行列式性质 (7), 得到

$$0 \leqslant \det((A-\alpha E_{11})\circ B)=\det(A\circ B)-\det\left(\begin{array}{c:c}\alpha b_{11} & a_{12}b_{12}\cdots a_{1n}b_{1n}\\ \hdashline 0 & A_{11}\circ B_{11}\end{array}\right)$$

$$=\det(A\circ B)-\alpha b_{11}\det(A_{11}\circ B_{11}).$$

再利用归纳假设 (3.6.3), 知

$$0\leqslant \det(A\circ B)-\alpha b_{11}\det A_{11}\cdot\prod_{i=2}^{n}b_{ii}=\det(A\circ B)^{-}\det A\prod_{i=1}^{n}b_{ii},$$

证毕.

对于矩阵的普通乘积, 我们有 $\det AB=\det A\cdot\det B$, 这里 A 与 B 都是方阵. 但对 Hadamard 乘积, 这不一定成立. 事实上, 我们有

推论 3.6.1　设 n 阶 Hermite 阵 $A>0, B>0$. 则

$$\det A\cdot\det B\leqslant\det(A\circ B)\leqslant\left(\prod_{i=1}^{n}a_{ii}\right)\left(\prod_{i=1}^{n}b_{ii}\right).$$

证明　对 B 应用 Hadamard 不等式, 再结合 (3.6.2), 便得到左边不等式. 另外, 由 $A>0, B>0$ 及定理 1.10.1(2), 知 $A\circ B>0$. 对 $A\circ B$ 应用 Hadamard 不等式 (3.3.1), 右边不等式得证. 定理证毕.

推论 3.6.2　若 Hermite 阵 $A>0$. 则 $\det(A\circ A^{-1})\geqslant 1$.

在推论 3.6.1 中, 令 $B=A^{-1}$, 即得欲证, 这个不等式再一次显示了 Hadamar 乘积与矩阵普通乘积的差别, 因为对后者, $\det AA^{-}=\det I=1$.

§3.7　Ostrowski-Taussky 不等式

对任一复方阵 A, 可分解为 $A=H+S$, 这里

$$H=\frac{1}{2}(A+A^*),\quad S=\frac{1}{2}(A-A^*)$$

分别为 Hermite 阵和斜 Hermite 阵 (即 $S^*=-S$). 我们称 H 为 A 的 Hermite 部分, 称 S 为 A 的斜 Hermite 部分, 下面的定理给出了 H 和 A 的行列式之间的关系.

定理 3.7.1　设 A 为 n 阶复方阵且 $H>0$. 则

$$\det H\leqslant|\det A|, \tag{3.7.1}$$

且等号成立 $\Longleftrightarrow$ A 是 Hermite 阵.

证明　容易验证 (3.7.1) 可用 H 和 S 表示为

$$|\det(I+H^{-1}S)|\geqslant 1. \tag{3.7.2}$$

用 $\lambda_i(B)(i=1,\cdots,n)$ 表示方阵 B 的特征值, 因为对相同阶的方阵 C 和 D, CD 与 DC 有相同的特征值, 故

$$\det(I+H^{-1}S)=\prod_{i=1}^{n}(1+\lambda_i(H^{-1}S))=\prod_{i=1}^{n}(1+\lambda_i(H^{-1/2}SH^{-1/2})).$$

注意到 $H^{-1/2}SH^{-1/2}$ 是斜 Hermite 阵, 它的特征值都为纯虚数 (见许以超 (1966), p.430, 定理 5), 所以

$$|1+\lambda_i(H^{-1/2}SH^{-1/2})|\geqslant 1,\quad i=1,\cdots,n,$$

这就证明了 (3.7.2). 其中等号成立 $\Longleftrightarrow$

$$\lambda_i(H^{-1/2}SH^{-1/2})=0,\quad i=1,\cdots,n. \tag{3.7.3}$$

因为斜 Hermite 阵是规范阵, 由定理 1.5.5. $H^{-1/2}SH^{-1/2}$ 必酉相似于对角阵, 结合 (3.7.3) 知, (3.7.2) 等号成立 $\Longleftrightarrow H^{-1/2}SH^{-1/2}=0\Longleftrightarrow S=0\Longleftrightarrow A$ 为 Hermite 阵. 证毕.

§3.8 华罗庚不等式

华罗庚(1955) 证明了下面的不等式, 并把它应用于复分析研究.

定理 3.8.1 设 A 和 B 为两个同阶复方阵, 且 $I-AA^*>0, I-BB^*>0$. 则

$$\det(I-AA^*)\det(I-BB^*)\leqslant|\det(I-AB^*)|^2,$$

等号成立 $\Longleftrightarrow A=B$.

实际上, 我们将证明下面更强的结果.

定理 3.8.2 在上面定理的条件下

$$\begin{aligned}&\det(I-AA^*)\cdot\det(I-BB^*)+|\det(A-B)|^2\\ \leqslant&|\det(I-AB^*)|^2,\end{aligned}$$

等号成立 $\Longleftrightarrow A=B$.

证明 因为

$$\begin{aligned}&(I-AB^*)(I-BB^*)^{-1}(I-AB^*)^*-(I-AA^*)\\ =&(I-BB^*)^{-1}-I-AB^*(I-BB^*)^{-1}-(I-BB^*)^{-1}BA^*\\ &+A(I+B^*(I-BB^*)^{-1}B)A^*,\end{aligned}$$

利用

$$(I-B^*B)^{-1}=I+B^*(I-BB^*)^{-1}B, \tag{3.8.1}$$

$$(I-BB^*)^{-1}=I+B(I-B^*B)^{-1}B^*, \tag{3.8.2}$$

上式可改写为

$$\begin{aligned}&(I-AB^*)(I-BB^*)^{-1}(I-AB^*)^*-(I-AA^*)\\ =&B(I-B^*B)^{-1}B^*-A(I-B^*B)^{-1}B^*-B(I-B^*B)^{-1}A^*\\ &+A(I-B^*B)^{-1}A^*\\ =&(A-B)(I-B^*B)^{-1}(A-B)^*.\end{aligned} \tag{3.8.3}$$

记

$$\begin{aligned}M_1&=I-AA^*,\\ M_2&=(A-B)(I-B^*B)^{-1}(A-B)^*,\end{aligned}$$

由 $I-BB^*>0$ 及 (3.8.1) 知 $M_2>0$. 从 (3.8.3) 得

$$(I-AB^*)(I-BB^*)^{-1}(I-AB^*)^*=M_1+M_2.$$

应用定理 3.2.1, 我们有

$$\begin{aligned}&\det[(I-AB^*)(I-BB^*)^{-1}(I-AB^*)^*]\\ \geqslant&\det M_1+\det M_2\\ =&\det(I-AA^*)+|\det(A-B)|^2(\det(I-B^*B))^{-1}.\end{aligned}$$

由定理 3.1.4, $\det(I-BB^*)=\det(I-B^*B)$, 代入上式, 命题得证.

关于这个不等式的进一步推广, 见 Marshall 和 Olkin(1979, p.235).

§3.9　Ky Fan 不等式

设 D 为一凸集, $f(\cdot)$ 为定义在 D 上一个实值函数, 若对任意 $x,y\in D$ 及 $0\leqslant\alpha\leqslant 1$, 总有

$$f(\alpha x+(1-\alpha)y)\leqslant\alpha f(x)+(1-\alpha)f(y),$$

则称 f 为 D 上的凸函数. 假如在上式中总是不等号 “$\geqslant$” 成立, 则称 f 为 D 上的凹函数.

若用 $M_{n\times n}$ 表示全体 n 阶正定 Hermite 阵, 则容易验证 $M_{n\times n}$ 为一凸集, 我们将证明 $-\ln\det(\cdot)$ 为 $M_{n\times n}$ 上的凸函数. 即对 $A>0,B>0,0\leqslant\alpha\leqslant 1$, 有

$$-\ln\det(\alpha A+(1-\alpha)B)\leqslant-\alpha\ln\det A-(1-\alpha)\ln\det B, \tag{3.9.1}$$

等价地

$$\det(\alpha A+(1-\alpha)B)\geqslant(\det A)^{\alpha}\cdot(\det B)^{1-\alpha}. \tag{3.9.2}$$

定理 3.9.1 设 A, B 为 n 阶正定 Hermite 阵, 则对一切 $0 \leqslant \alpha \leqslant 1$, (3.9.2) 成立且等号成立 $\Longleftrightarrow A = B$ 或 $\alpha = 0$ 或 $\alpha = 1$.

证明 从 (3.9.1) 知, 问题归结为证明

$$\text{lndet}(\alpha A + (1-\alpha)B) \geqslant \alpha \text{lndet} A + (1-\alpha)\text{lndet} B. \tag{3.9.3}$$

因为 $A > 0, B > 0$, 应用定理 1.4.4, 存在可逆阵 P, 使得

$$\begin{aligned} A &= PP^*, \\ B &= P\Lambda P^*, \end{aligned}$$

这里 $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n), \lambda_i > 0, i = 1, \cdots, n$. 于是 (3.9.3) 的左边可改写为

$$\begin{aligned} &\text{lndet}(\alpha A + (1-\alpha)B) \\ =& \text{lndet} P(\alpha I + (1-\alpha)\Lambda)P^* \\ =& \text{lndet}(PP^*) + \text{lndet}(\alpha I + (1-\alpha)\Lambda) \\ =& \text{lndet} A + \text{lndet}(\alpha I + (1-\alpha)\Lambda). \end{aligned} \tag{3.9.4}$$

另一方面, (3.9.3) 的右边

$$\begin{aligned} &\alpha \text{lndet} A + (1-\alpha)\text{lndet} B \\ =& \alpha \text{lndet} A + (1-\alpha)[\text{lndet} PP^* + \text{lndet}\Lambda] \\ =& \text{lndet} A + (1-\alpha)\text{lndet}\Lambda. \end{aligned} \tag{3.9.5}$$

比较 (3.9.4) 和 (3.9.5) 知, 我们只需证明

$$\text{lndet}(\alpha I + (1-\alpha)\Lambda) \geqslant (1-\alpha)\text{lndet}\Lambda. \tag{3.9.6}$$

因为对数函数是严格凹函数, 故

$$\begin{aligned} &\text{lndet}(\alpha I + (1-\alpha)\Lambda) \\ =& \ln \prod_{i=1}^{n} (\alpha + (1-\alpha)\lambda_i) \\ =& \sum_{i=1}^{n} \ln(\alpha + (1-\alpha)\lambda_i) \\ \geqslant& \sum_{i=1}^{n} \alpha \ln 1 + (1-\alpha)\ln\lambda_i \\ =& (1-\alpha)\sum_{i=1}^{n} \ln\lambda_i = (1-\alpha)\text{In}\prod_{i=1}^{n} \lambda_i \\ =& (1-\alpha)\text{lndet}\Lambda. \end{aligned} \tag{3.9.7}$$

(3.9.6) 得证. 由 $\ln x$ 严格凹性, 从 (3.9.7) 知等号成立的充要条件为 $\alpha = 0$ 或 1 或 $\lambda_i = 1, i = 1, \cdots, n$. 这等价于 $\alpha = 0$ 或 1 或 $B = A$. 定理证毕.

若记 $\lambda_1(C) \geqslant \cdots \geqslant \lambda_n(C)$ 为正定 Hermite 阵 C 的特征值, 则 (3.9.2) 的另一种等价形式为

$$\prod_{i=1}^{n} \lambda_i(\alpha A + (1-\alpha)B) \geqslant \prod_{i=1}^{n} (\lambda_i(A))^{\alpha} (\lambda_i(B))^{1-\alpha}, \quad 0 \leqslant \alpha \leqslant 1.$$

推论 3.9.1　在定理 3.9.1 条件下, 对任意的 $1 \leqslant k \leqslant n$, 有

$$\prod_{i=k}^{n} \lambda_i(\alpha A + (1-\alpha)B) \geqslant \prod_{i=k}^{n} (\lambda_i(A))^{\alpha} (\lambda_i(B))^{1-\alpha}, \quad 0 \leqslant \alpha \leqslant 1.$$

证明　记 $\alpha A + (1-\alpha)B$ 的后 $n-k+1$ 个特征值对应的标准正交化特征向量组成的矩阵为 Q, 则 $Q^*Q = I_{n-k+1}$. 利用 (3.9.2), 我们有

$$\begin{aligned}
&\prod_{i=k}^{n} \lambda_i(\alpha A + (1-\alpha)B) \\
&= \det Q^*(\alpha A + (1-\alpha)B)Q \\
&= \det(\alpha Q^*AQ + (1-\alpha)Q^*BQ) \\
&\geqslant (\det Q^*AQ)^{\alpha} \cdot (\det Q^*BQ)^{1-\alpha} \\
&= \left(\prod_{i=1}^{n=k+1} \lambda_i(Q^*AQ)\right)^{\alpha} \left(\prod_{i=1}^{n-k+1} \lambda_i(Q^*BQ)\right)^{1-\alpha}.
\end{aligned} \tag{3.9.8}$$

应用 Poincaré 分离定理 (见定理 4.6.1), 得

$$\lambda_i(Q^*AQ) \geqslant \lambda_{k+i-1}(A), \quad i = 1, \cdots, n-k+1,$$

$$\lambda_i(Q^*BQ) \geqslant \lambda_{k+i-1}(B), \quad i = 1, \cdots, n-k+1,$$

代入 (3.9.8), 即可得证.

§3.10　Lavoie 不等式

设 A 为 $m \times n$ 矩阵, A^+ 为其 Moore-Penrosé广义逆. 因为 $AA^+ = A(A^*A)^+A^*$, 故 AA^+ 是正交投影阵, 于是

$$\det AA^+ = \begin{cases} 1, & r(A) = m, \\ 0, & r(A) < m. \end{cases}$$

Lavoie(1980) 研究了推广情形, 他的结果概括为下面的定理.

定理 3.10.1 (Lavoie) 设 A 为 $m\times n$ 复矩阵, 将其分块为

$$A=\begin{pmatrix}A_1\\A_2\\\vdots\\A_k\end{pmatrix},\quad A_i \text{ 为 } m_i\times n \text{ 复矩阵},\quad \sum_{i=1}^k m_i=m.$$

记 $B=(A_1^+,\cdots,A_k^+)$, 则

$$0\leqslant \det AB\leqslant 1,\tag{3.10.1}$$

且左边等号成立 $\Longleftrightarrow r(A)<m$; 而右边等号成立 $\Longleftrightarrow A_iA_j^*=0$, 对一切 $i\neq j$ 成立.

证明 若 $r(A)<m$, 显然 $\det AB=0$. 以下假设 $r(A)=m$, 此时 $r(A_i)=m_i$, 由定理 1.7.10(3) 知

$$A_i^+=A_i^*(A_iA_i^*)^{-1},\quad i=1,\cdots,k.$$

若记

$$D=\begin{pmatrix}(A_1A_1^*)^{-1}&0&\cdots&0\\0&(A_2A_2^*)^{-1}&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&(A_kA_k^*)^{-1}\end{pmatrix},$$

则 $B=A^*D$,

$$\det AB=\det AA^*\det D=\frac{\det AA^*}{\prod\limits_{i=1}^k\det(A_iA_i^*)}.\tag{3.10.2}$$

因为 $AA^*>0, D>0$, 所以 $\det AB>0$, 从 (3.10.2) 知, 我们只需证明

$$\det AA^*\leqslant\prod_{i=1}^k\det A_iA_i^*.\tag{3.10.3}$$

由 Fischer 定理 (3.4.1) 立即可得 (3.10.3), 且等号成立 $\Longleftrightarrow A_iA_j^*=0$, 对一切 $i\neq j$. 定理证毕.

推论 3.10.1 在定理 3.10.1 条件下,

$$r(AB)=r(AA^*)=r(A).$$

证明 采用定理 3.10.1 的记号, 我们有 $AB=AA^*D$. 因为 D 是可逆阵, 故 $r(AB)=r(AA^*)=r(A)$.

上面的定理及推论的证明是本书作者给出的, Styan(1981) 给出了另外一种证明, Khatri(1983) 把 (3.10.1) 推广到 $B=(B_1,\cdots,B_k), B_i=A_i^{(1,4)}(i=1,\cdots,k)$ 的情形.

§3.11　其　　他

本节要讨论的不等式有两类, 第一类即定理 3.11.1, 它给出了任意复方阵行列式的模的上下界, 这些界只包含了该方阵的元素. 而第二类即最后的两个定理, 它们给出的是矩阵积的行列式的界, 这些界只包含矩阵的特征值.

定理 3.11.1　设 $A=(a_{ij})$ 为 $n\times n$ 复方阵, 记

$$S_i=\sum_{j=i+1}^{n}|a_{ij}|,\quad i=1,\cdots,n,$$

若

$$|a_{ii}|>\sum_{j\neq i}|a_{ij}|,\quad i=1,\cdots,n, \tag{3.11.1}$$

则

$$0<\prod_{i=1}^{n}(|a_{ii}|-S_i)\leqslant|\det A|\leqslant\prod_{i=1}^{n}(|a_{ii}|+S_i), \tag{3.11.2}$$

且当 A 为下三角阵时, 等号成立.

证明　我们首先证明, 在条件 (3.11.1) 下, $\det A\neq 0$. 用反证法, 假设 $\det A=0$, 则线性方程组 $Ax=0$ 有非零解 $\widetilde{x}=(\widetilde{x}_1,\cdots,\widetilde{x}_n)'$. 记

$$|\widetilde{x}_k|=\max_i\{|\widetilde{x}_i|\}.$$

从 $A\widetilde{x}=0$ 得

$$|a_{kk}\widetilde{x}_k|=|\sum_{j\neq k}a_{kj}\widetilde{x}_j|\leqslant\sum_{j\neq k}|a_{kj}||\widetilde{x}_k|=|\widetilde{x}_k|\sum_{j\neq k}|a_{kj}|,$$

即

$$|a_{kk}|\leqslant\sum_{j\neq k}|a_{kj}|. \tag{3.11.3}$$

这与 (3.11.1) 相矛盾.

记

$$A_1=\begin{pmatrix}a_{22}&\cdots&a_{2n}\\ \vdots&&\vdots\\ a_{n2}&\cdots&a_{nn}\end{pmatrix},\qquad y=\begin{pmatrix}x_2\\ \vdots\\ x_n\end{pmatrix},\qquad b=\begin{pmatrix}a_{21}\\ \vdots\\ a_{n1}\end{pmatrix},$$

和上面同样的道理, $\det A_1\neq 0$. 于是线性方程组 $A_1y=-b$ 有唯一解 $y^{(1)}=(x_2^{(1)},\cdots,x_n^{(1)})$. 设

$$|x_k^{(1)}|=\max_{1\leqslant i\leqslant n-1}\{|x_i^{(1)}|\},$$

类似于 (3.11.3) 的证明, 可证明

$$|a_{kk}| \leqslant \sum_{\substack{j=2 \\ j\neq k}}^{n} |a_{kj}| + \frac{|a_{k1}|}{|x_k^{(1)}|}. \tag{3.11.4}$$

结合 (3.11.1) 推得

$$|x_k^{(1)}| < 1. \tag{3.11.5}$$

记 $c_1 = (a_{12}, \cdots, a_{1n})'$, 则因

$$\begin{pmatrix} a_{11} & c_1' \\ b & A_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y^{(1)} & I_{n-1} \end{pmatrix} = \begin{pmatrix} a_{11} + c_j' y^{(1)} & c_1' \\ 0 & A_1 \end{pmatrix},$$

故有

$$\det A = (a_{11} + c_j' y^{(1)}) \det A_1.$$

而从 (3.11.5), 我们有

$$|c_j' y^{(1)}| < \sum_{j=2}^{n} |a_{1j}|.$$

用完全相同的方法可以证明

$$\det A_1 = (a_{22} + c_2' y^{(2)}) \det A_2,$$
$$|c_2' y^{(2)}| < \sum_{j=3}^{n} |a_{2j}|,$$

其中

$$A_2 = \begin{pmatrix} a_{33} & \cdots & a_{3n} \\ \vdots & & \vdots \\ a_{n3} & \cdots & a_{nn} \end{pmatrix}, \qquad c_2 = \begin{pmatrix} a_{23} \\ \vdots \\ a_{2n} \end{pmatrix},$$

$$y^{(2)} = (x_3^{(2)}, \cdots, x_n^{(2)}), \quad |x_j^{(2)}| < 1, \quad j = 3, \cdots, n.$$

继续这样做下去, 最后我们得到

$$\det A = \prod_{i=1}^{n} (a_{ii} + c_i' y^{(i)}),$$
$$|c_i' y^{(i)}| < \sum_{i=j+1}^{n} |a_{ij}|, \quad i = 1, \cdots, n.$$

再应用初等不等式 $|a| - |b| \leqslant |a+b| \leqslant |a| + |b|$, 我们有

$$0 < \prod_{i=1}^{n} \left(|a_{ii}| - \sum_{j=i+1}^{n} |a_{ij}| \right) \leqslant |\det A| \leqslant \prod_{i=1}^{n} \left(|a_{ii}| + \sum_{j=i+1}^{n} |a_{ij}| \right).$$

显然, 当 A 为下三角阵时, $S_i = 0, i = 1, \cdots, n$ 所以此时等号成立. 定理证毕.

下面我们叙述第二类的不等式, 它们的证明将在下一章给出.

定理 3.11.2　设 A 为 $n \times n$ 半正定 Hermite 阵, 则对一切满足 $X^*X = I_k$ 的 $n \times k$ 矩阵 X, 有

$$\prod_{i=n-k+1}^{n} \lambda_i \leqslant \det X^*AX \leqslant \prod_{i=1}^{k} \lambda_i,$$

这里 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值.

证明需要利用 Poincaré 定理, 见推论 4.6.1.

定理 3.11.3　设 A 为 $n \times n$ 正定 Hermite 阵, 则对一切满足 $X^*X = I_k$ 的 $n \times k$ 矩阵 X, 有

$$\det(X^*AX)\det(X^*A^{-1}X) \leqslant \prod_{i=1}^{k} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}},$$

这里 $\lambda_i \geqslant \cdots \geqslant \lambda_n (n > 2k)$ 为 A 的特征值.

证明见定理 4.10.3.

第4章 特 征 值

本章研究特征值的各种不等式，在前两节我们先证明两个重要定理：Rayleigh-Ritz 定理和 Courant-Fischer 定理，它们通过 Rayleigh-Ritz 商的各种极值来表征 Hermite 阵的特征值；作为这两个定理的应用，接下来的 §4.3, §4.4 和 §4.5 分别导出了镶边矩阵特征值的交叉性质和两个 Hermite 阵之和的特征值的一些不等式以及 Sturm 定理；在 §4.6 应用 Sturm 定理，我们证明了应用广泛的 Poincaré 定理及其他一些有关矩阵积的特征值的不等式. 对一般的复或实方阵的特征值，一般说来我们所知较少；§4.7 证明了 Schur 定理，Hirsch 定理和 Bendixson 定理，它们给出了复方阵或实方阵的特征值或实部或虚部的界；在本章最后一节，我们证明了 Wielandt 不等式，它可以看作 Cauchy-Schwarz 不等式 (1.9.2) 的一种改进.

鉴于特征值的重要性，本章的结果在工程技术、物理、数值计算、概率统计及其他数学分支有着重要应用，特别，关于在数理统计中的一些应用，读者可参阅王松桂 (1985) 以及 Wang Songgui 等 (1990).

§4.1 Rayleigh-Ritz 定理

众所周知，对于一般的方阵，其特征值是特征多项式的根. 若进一步假设它是 Hermite 阵，那么此时它的特征值是实数，并且可表示为Rayleigh-Ritz 商 x^*Ax/x^*x 的极值.

我们将用 $(\lambda_1 \geqslant \cdots \geqslant \lambda_n)$ 表示方阵 A 的特征值，$\varphi_1, \cdots, \varphi_n$ 表示对应的标准正交化特征向量. 在必要时，也采用 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ 表示方阵 A 的特征值.

定理 4.1.1 (Rayleigh-Ritz) 设 A 为 $n \times n$ 的 Hermite 阵. 则

$$\lambda_1 = \max_{x \neq 0} \frac{x^*Ax}{x^*x}, \tag{4.1.1}$$

$$\lambda_n = \min_{x \neq 0} \frac{x^*Ax}{x^*x}. \tag{4.1.2}$$

证明 记 $\Phi = (\varphi_1, \cdots, \varphi_n), \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 因为 $\varphi_1, \cdots, \varphi_n$ 构成 C^n 的基，所以对任给的 $x \in C^n$，存在向量 t，使得 $x = \Phi t$. 故

$$\frac{x^*Ax}{x^*x} = \frac{t^*\Lambda t}{t^*t} = \sum_{i=1}^{n} \lambda_i \frac{|t_i|^2}{t^*t} \leqslant \lambda_1 \sum_{i=1}^{n} \frac{|t_i|^2}{t^*t} = \lambda_1$$

达到最大值 $\Longleftrightarrow t_i=0, i\geqslant 2$, 且 $t_1\neq 0 \Longleftrightarrow x=c\varphi_1$, 对某个数 c. 这就证明了 (4.1.1). 类似地可证 (4.1.2).

注 1　(4.1.1) 和 (4.1.2) 可以合并为如下不等式

$$\lambda_n x^*x \leqslant x^*Ax \leqslant \lambda_1 x^*x. \tag{4.1.3}$$

推论 4.1.1　设 A 为 $n\times n$ Hermite 阵, $x\neq 0$. 记

$$a=\frac{x^*Ax}{x^*x}.$$

则至少存在 A 的两个特征值 λ 和 μ, 使得

$$-\infty<\lambda\leqslant a, a\leqslant \mu<\infty.$$

推论 4.1.2　设 $A=(a_{ij})$ 为 $n\times n$ Hermite 阵, 则

$$\lambda_n\leqslant a_{ii}\leqslant \lambda_1,\quad i=1,\cdots,n.$$

证明　在 (4.1.3) 中取 $x=(0,0,1,0,\cdots,0)'$, 这里 1 为 x 的第 i 个元素, 即得本推论.

注 2　定理 4.1.1 的结论还可以改写为如下形式

$$\lambda_1=\max_{x^*x=1} x^*Ax=\varphi_1^*A\varphi_1, \tag{4.1.4}$$

$$\lambda_n=\min_{x^*x=1} x^*Ax=\varphi_n^*A\varphi_n. \tag{4.1.5}$$

从几何上, 这两个结果可解释为: A 的最大和最小特征值分别是二次型 x^*Ax 在单位球 $x^*x=1$ 上的最大值和最小值. 对于其他特征值, 我们也有类似结果, 但此时 x 必须限制在某一子空间上.

定理 4.1.2　设 A 为 $n\times n$ Hermite 阵, 则

$$\lambda_k=\max_{\substack{x\neq 0\\ x\perp\varphi_j(j=1,\cdots,k-1)}}\frac{x^*Ax}{x^*x} \tag{4.1.6}$$

$$\lambda_{n-k}=\min_{\substack{x\neq 0\\ x\perp\varphi_j(j=n-k+1,\cdots,n)}}\frac{x^*Ax}{x^*x}. \tag{4.1.7}$$

证明　在定理 4.1.1 证明中, 分别用 $\mathscr{M}(\varphi_k,\cdots,\varphi_n)$ 和 $\mathscr{M}(\varphi_1,\cdots,\varphi_{n-k})$ 代替 C^n, 即得 (4.1.6) 和 (4.1.7).

下面的定理是将 Raylelgh-Ritz 定理作进一步推广, 给出了两个二次型商 x^*Ax/x^*Bx 的极值.

定理 4.1.3 设 A 和 B 皆为 $n\times n$ Hermite 阵, 且 $B>0$. 记 $\mu_1\geqslant\cdots\geqslant\mu_n$ 为 A 关于 B 的相对特征值, 即 μ_i 为 $\det(A-\mu B)=0$ 的根. 则

$$\mu_1=\max_{x\neq 0}\frac{x^*Ax}{x^*Bx}, \tag{4.1.8}$$

$$\mu_n=\min_{x\neq 0}\frac{x^*Ax}{x^*Bx}. \tag{4.1.9}$$

证明 记 $y=B^{1/2}x$, 应用定理 4.1.1 得

$$\begin{aligned}\max_{x\neq 0}\frac{x^*Ax}{x^*Bx}&=\max_{y\neq 0}\frac{x^*B^{-1/2}AB^{-1/2}y}{y^*y}=\lambda_1(B^{-1/2}AB^{-1/2})\\&=\lambda_1(AB^{-1})=\mu_1,\end{aligned}$$

其中我们应用了特征值的性质: CD 和 DC 有相同特征值 (见定理 1.2.7). 类似方法可证 (4.1.9). 定理证毕.

注 3 (4.1.8) 和 (4.1.9) 可以改写为如下不等式形式

$$\mu_n\leqslant\frac{x^*Ax}{x^*Bx}\leqslant\mu_1,$$

或

$$\mu_n x^*Bx\leqslant x^*Ax\leqslant\mu_1 x^*Bx.$$

§4.2 Courant-Fischer 定理

在上节, 我们把 Hermite 阵的特征值表为 Rayleigh-Ritz 商的极值, 如 (4.1.6) 和 (4.1.7). 除了最大和最小特征值外, 这些特征值都是 Rayleigh-Ritz 商的约束极值, 且约束条件与该方阵的特阵向量有关, 有时这是不方便的. 下面的 Courant-Fischer 定理克服了这一缺陷.

定理 4.2.1 (Courant-Fischer) 设 A 为 $n\times n$ Hermite 阵, $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为其特征值, $\varphi_1,\cdots,\varphi_n$ 为对应的标准正交化特征向量. 设 B 为 $n\times k$ 矩阵. 则

(1) $$\min_B\max_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}=\lambda_{k+1}, \tag{4.2.1}$$

且当 $B=(\varphi_1,\cdots,\varphi_k)$ 时上式达到 λ_{k+1};

(2) $$\max_B\min_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}=\lambda_{n-k}, \tag{4.2.2}$$

且当 $B=(\varphi_{n-k+1},\cdots,\varphi_n)$ 时上式达到 λ_{n-k}.

证明　(1) 记 $\Phi=(\varphi_1,\cdots,\varphi_n), x=\Phi y$, 则

$$\begin{aligned}\max_{B^*x=0}\frac{x^*Ax}{x^*x}=\max_{C^*y=0}\frac{y^*\Lambda y}{y^*y}&\\ &\geqslant \max_{C^*\binom{y_1}{0}=0}\frac{y_1^*\Lambda_1y_1}{y_1^*y_1}\\ &\geqslant \min_{C^*\binom{y_1}{0}=0}\frac{y_1^*\Lambda_1y_1}{y_1^*y_1}\\ &\geqslant \min_{y_1\neq 0}\frac{y_1^*\Lambda_1y_1}{y_1^*y_1}=\lambda_{k+1}.\end{aligned}$$

这里 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n), C=\Phi^*B, \Lambda_1=\mathrm{diag}(\lambda_1,\cdots,\lambda_{k+1}), y'=(y_1',y_2'), y$ 为 $(k+1)\times 1$ 向量.
于是

$$\min_B\max_{B^*x=0}\frac{x^*Ax}{x^*x}\geqslant\lambda_{k+1}.$$

由定理 4.1.2 知, 当 $B=(\varphi_1,\cdots,\varphi_k)$ 时, 上式取等号.

(2) 采用与上段完全相同的记号. 并记

$$\Lambda_2=\mathrm{diag}(\lambda_{n-k},\cdots,\lambda_n),$$

$$y'=(y_1',y_2'), y_2 \text{ 为 } (n-k)\times 1 \text{ 向量}.$$

则有

$$\begin{aligned}\min_{B^*x=0}\frac{x^*Ax}{x^*x}=\min_{C^*y=0}\frac{y^*\Lambda y}{y^*y}&\\ &\leqslant \min_{C^*\binom{0}{y_2}=0}\frac{y_2^*\Lambda_2y_2}{y_2^*y_2}\\ &\leqslant \max_{C^*\binom{0}{y_2}=0}\frac{y_2^*\Lambda_2y_2}{y_2^*y_2}\leqslant\max\frac{y_2^*\Lambda_2y_2}{y_2^*y_2}=\lambda_{n-k},\end{aligned}$$

于是

$$\max_B\min_{B^*x=0}\frac{x^*Ax}{x^*x}\leqslant\lambda_{n-k}.$$

根据定理 4.1.2 知, 当 $B=(\varphi_{n-k+1},\cdots,\varphi_n)$ 时, 上式取等号. 定理证毕.

注　从 (4.2.1) 和 (4.2.2) 我们看到, Hermite 阵 A 的任一特征值都可表为 Rayleigh-Ritz 商 x^*Ax/x^*x 的约束极值. 而约束条件与 A 的特征向量无关, 而是涉及一般的矩阵 B. 正是由于这一点, Courant-Fischer 定理有着极为广泛的应用. 后面几节我们将应用它导出许多特征值不等式.

定理 4.2.2 设 A 为 $n \times n$ Hermite 阵, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为其特征值, $\varphi_1, \cdots, \varphi_n$ 为对应的标准正交化特征向量, 设 $x_1, \cdots, x_k$ 满足 $x_i^* x_j = 0, i \neq j$. 则

$$(1)\ \max_{x_1,\cdots,x_k} \sum_{i=1}^{k} \frac{x_i^* A x_i}{x_i^* x_i} = \sum_{i=1}^{k} \lambda_i, \tag{4.2.3}$$

且当 $x_i \propto \varphi_i, i = 1, \cdots, k$ 时达到最大值.

$$(2)\ \min_{x_1\cdots x_k} \prod_{i=1}^{k} \frac{x_i^* A x_i}{x_i^* x_i} = \prod_{i=n-k+1}^{n} \lambda_i, \tag{4.2.4}$$

且当 $x_i \propto \varphi_{n-k+i}, i = 1, \cdots, k$ 时达到最小值.

证明 (1) 不妨设 $x_i^* x_i = 1, i = 1, \cdots, k$. 因 $\varphi_1, \cdots, \varphi_n$ 为 C^n 的标准正交基, 记 $\Phi = (\varphi_1, \cdots, \varphi_n)$, 则 Φ 为一酉阵, 存在向量 $c_1, \cdots, c_k$, 使得 $x_i = \Phi c_i, i = 1, \cdots, k$, 其中 $c_i = (c_{1i}, \cdots, c_{ni})'$. 记 $C = (c_1, \cdots, c_k)$. 则由

$$x_i^* x_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

可推得

$$c_i^* c_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

即

$$C^* C = I_k. \tag{4.2.5}$$

故可将 C 扩充为一酉阵 $B = (C \vdots D)$. 由此容易证明

$$\begin{cases} \alpha_j \triangleq \sum_{i=1}^{k} |c_{ji}|^2 \leqslant 1, \quad j = 1, \cdots, n, \\ \sum_{j=1}^{n} \alpha_j = k. \end{cases} \tag{4.2.6}$$

另一方面

$$\begin{aligned} \sum_{i=1}^{k} \frac{x_i^* A x_i}{x_i^* x_i} &= \sum_{i=1}^{k} x_i^* A x_i = \sum_{i=1}^{k} \sum_{j=1}^{n} |c_{ji}|^2 \lambda_j = \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{k} |c_{ji}|^2 \\ &= \sum_{j=1}^{n} \alpha_j \lambda_j, \end{aligned} \tag{4.2.7}$$

为使 (4.2.7) 极大化, 从 (4.2.6) 容易看出, 应选择 c_{ij}, 使得

$$\alpha_1 = \cdots = \alpha_k = 1, \alpha_{k+1} = \cdots = \alpha_n = 0,$$

这样的 c_{ij} 是存在的. 例如, 对 $i=1,\cdots,k$, 取

$$c_{ji}=\begin{cases}1, & j=i,\\ 0, & j\neq i.\end{cases}\qquad j=1,\cdots,n,$$

这相当于取 $x_i=\varphi_i, i=1,\cdots,k$. 得证.

(2) 和上段一样, 不妨设 $x_i^*x_i=1, x_i=\Phi c_i, i=1,\cdots,k$. 则 $c_i^*c_i=1, i=1,\cdots,k$. 记 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, 则 (4.2.4) 变形为

$$\min_{c_1,\cdots,c_k}\prod_{i=1}^{k}c_i^*\Lambda c_i=\prod_{i=1}^{k}\lambda_i.$$

应用推广的算术平均与几何平均不等式 (见附录 1 之 1(5)), 我们有

$$c_i^*\Lambda c_i\geqslant\lambda_1^{|c_{1i}|^2}\cdots\lambda_2^{|c_{ni}|^2},\quad i=1,\cdots,k.$$

于是

$$\prod_{i=1}^{k}c_i^*\Lambda c_i\geqslant\lambda_1^{\alpha_1}\cdots\lambda_n^{\alpha_n},\tag{4.2.8}$$

这里 α_j 定义同 (4.2.5). 易见, 为使 (4.2.8) 右端极小化, 应选 c_{ij}, 使得

$$\alpha_j=\begin{cases}0, & j\leqslant n-k,\\ 1, & j\geqslant n-k+1.\end{cases}$$

类似于 (1) 中的证明, 这相当于 $x_i=\varphi_i, i=n-k+1,\cdots,n$. 证毕.

类似于定理 4.1.3, 我们能够证明如下推论.

推论 4.2.1 设 A、B 皆为 $n\times n$ Hermite 阵, $B>0$. 记 $\mu_1\geqslant\cdots\geqslant\mu_n$ 为 A 关于 B 的相对特征值. 假设 $x_1,\cdots,x_k$ 满足 $x_i^*Cx_j=0, i\neq j$, 则

$$\max_{x_1,\cdots,x_k}\sum_{i=1}^{k}\frac{x_i^*Ax_i}{x_i^*Cx_i}=\sum_{i=1}^{k}\mu_i,$$

$$\min_{x_1,\cdots,x_k}\prod_{i=1}^{k}\frac{x_i^*Ax_i}{x_i^*Cx_i}=\prod_{i=n=k+1}^{n}\mu_i.$$

证明留作练习, 请读者自己去完成.

在结束这一节的时候, 作为 Courant-Fischer 定理的一个应用, 我们导出复方阵 A 的奇异值与 A 的 Hermite 部分 $H=(A+A^*)/2$ 的特征值之间的一个不等式.

定理 4.2.3 设 A 为 $n\times n$ 复方阵, $\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$ 为其奇异值, $\lambda_1(H)\geqslant\cdots\geqslant\lambda_n(H)$ 为 H 的特征值. 则

$$\lambda_i(H)\leqslant\sigma_i(A),\quad i=1,\cdots,n.$$

证明 设 A 的奇异值分解为

$$A = U\Delta V^*, \tag{4.2.9}$$

其中 U 和 V 均为酉阵, $\Delta = \mathrm{diag}(\sigma_1(A), \cdots, \sigma_n(A))$, 将 (4.2.9) 改写为

$$A = UV^*V\Delta V^* \triangleq PQ, \tag{4.2.10}$$

其中 $P = UV^*$ 仍为酉阵, $Q = V\Delta V^* \geqslant 0$. 对任一 $n \times 1$ 单位向量 x, 利用 (4.2.10) 及 Cauchy-Schwarz 不等式, 得

$$x^*Hx = \mathrm{Re}x^*Ax = \mathrm{Re}x^*PQx \leqslant |x^*PQx| \leqslant (x^*Q^2x)^{1/2}.$$

应用定理 4.2.1, 我们有

$$\begin{aligned}\lambda_i(H) &= \min_{\substack{B\\(i-1)\times n}} \max_{\substack{Bx=0\\\|x\|=1}} x^*Hx \leqslant \min_{\substack{B\\(i-1)\times n}} \max_{\substack{Bx=0\\\|x\|=1}} (x^*Q^2x)^{1/2}\\ &= (\lambda_i(Q^2))^{1/2} = \lambda_i^{1/2}(A^*A) = \sigma_i(A), \quad i = 1, \cdots, n.\end{aligned}$$

证毕.

§4.3 镶边矩阵的特征值

设 A 为 $n \times n$ 方阵, a 为一数, b 为 $n \times 1$ 向量, 我们称矩阵

$$B = \begin{pmatrix} A & b \\ b^* & a \end{pmatrix} \tag{4.3.1}$$

为镶边矩阵, 镶边矩阵 B 的特征值与原来矩阵 A 的特征值有密切关系. 本节将应用 Courant-Fischer 定理建立这方面的一些重要不等式.

定理 4.3.1 设 A 为 Hermite 阵, a 为一实数. B 是由 (4.3.1) 定义的镶边矩阵, 记 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 和 $\mu_1 \geqslant \cdots \geqslant \mu_{n+1}$ 分别为 A 和 B 的特征值. 则

$$\mu_1 \geqslant \lambda_1 \geqslant \mu_2 \geqslant \cdots \geqslant \lambda_{n-1} \geqslant \mu_n \geqslant \lambda_n \geqslant \mu_{n+1}. \tag{4.3.2}$$

证明 我们需证明

$$\mu_{i+1} \leqslant \lambda_i \leqslant \mu_i, \quad i = 1, \cdots, n.$$

设 $y' = (x', t)$ 和 $z_i' = (u_i', v)$ 为 $(n+1)\times 1$ 向量, 而其中的 x 和 $u_i(i=1,\cdots,n)$ 为 $n\times 1$ 向量. 应用 Courant-Fischer 定理, 有

$$\begin{aligned}
\mu_i &= \min_{z_1,\cdots,z_{i-1}} \max_{\substack{y\perp z_j \\ j=1,\cdots,i-1}} \frac{y^*By}{y^*y} \\
&\geqslant \min_{z_1,\cdots,z_{i-1}} \max_{\substack{y\perp z_j \\ j=1,\cdots,i-1 \\ t=0}} \frac{y^*By}{y^*y} \\
&= \min_{u_1,\cdots,u_{i-1}} \max_{\substack{x\perp u_j \\ j=1,\cdots,i-1}} \frac{x^*Ax}{x^*x} = \lambda_i, \quad i=2,\cdots,n.
\end{aligned}$$

而应用 Rayleigh-Ritz 定理

$$\mu_1 = \max_y \frac{y^*By}{y^*y} \geqslant \max_{\substack{y \\ t=0}} \frac{y^*By}{y^*y} = \max_x \frac{x^*Ax}{x^*x} = \lambda_1.$$

另一方面, 同样由 Courant-Fischer 定理, 有

$$\begin{aligned}
\mu_{i+1} &= \max_{z_1,\cdots,z_{n-i-1}} \min_{\substack{y\perp z_j \\ j=1,\cdots,n-i-1}} \frac{y^*By}{y^*y} \\
&\leqslant \max_{z_1,\cdots,z_{n-i-1}} \min_{\substack{y\perp z_j \\ j=1,\cdots,n-i-1 \\ t=0}} \frac{y^*By}{y^*y} \\
&= \max_{u_1,\cdots,u_{n-i-1}} \min_{\substack{x\perp u_j \\ j=1,\cdots,n-i-1}} \frac{x^*Ax}{x^*x} = \lambda_i, \quad i=1,\cdots,n-1,
\end{aligned}$$

应用 Rayleigh-Ritz 定理, 得

$$\mu_{n+1} = \min_y \frac{y^*By}{y^*y} \leqslant \min_{\substack{y \\ t=0}} \frac{y^*By}{y^*y} = \min_x \frac{x^*Ax}{x^*x} = \lambda_n.$$

定理证毕.

这个定理表明, 镶边矩阵与原来矩阵的特征值满足交叉关系 (4.3.2). 现在考虑它的反问题. 如果给定了两组实数 $\{\lambda_i\}_{i=1}^n$ 和 $\{\mu_j\}_{j=1}^{n+1}$, 它们满足关系 (4.3.2). 问题是能否找到 Hermite 阵 A 及镶边矩阵 B, 使得它们的特征值分别为给定的 $\{\lambda_i\}_{i=1}^n$ 和 $\{\mu_j\}_{j=1}^{n+1}$. 下面的定理给了这个问题一个肯定回答.

定理 4.3.2 设 n 为一正整数, 给定两组实数 $\{\lambda_i\}_{i=1}^n$ 和 $\{\mu_j\}_{j=1}^{n+1}$, 假设它们满足 (4.3.2). 记 $\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, 则存在实数 a 及 $n\times 1$ 实向量 b, 使得 $\{\mu_j\}_{j=1}^{n+1}$ 为实对称阵

$$B = \begin{pmatrix} \Lambda & b \\ b' & a \end{pmatrix}$$

的特征值.

证明 首先定义

$$a=\sum_{j=1}^{n+1}\mu_j-\sum_{i=1}^{n}\lambda_i. \tag{4.3.3}$$

方阵 B 的特征多项式 $\varphi_B(t)$ 为

$$\begin{aligned}\varphi_B(t)&=\det(tI_{n+1}-B)\\&=\begin{vmatrix}tI_n-\Lambda & -b\\ -b' & t-a\end{vmatrix}\\&=\left|\begin{pmatrix}I_n & 0\\ ((tI_n-\Lambda)^{-1}b)' & 1\end{pmatrix}\begin{pmatrix}tI_n-\Lambda & -b\\ -b' & t-a\end{pmatrix}\begin{pmatrix}I_n & (tI_n-\Lambda)^{-1}b\\ 0 & 1\end{pmatrix}\right|\\&=[(t-a)-b'(tI_n-\Lambda)^{-1}b]\det(tI_n-\Lambda)\\&=\left[(t-a)-\sum_{i=1}^{n}\frac{b_i^2}{t-\lambda_i}\right]\prod_{i=1}^{n}(t-\lambda_i).\end{aligned} \tag{4.3.4}$$

因为 (4.3.3) 已把 a 选定, 以下我们只需证明存在 $b=(b_1,\cdots,b_n)'$, 使得 $\varphi_B(\mu_j)=0, j=1,\cdots,n+1$.

记

$$\boldsymbol{\Psi}_1(t)=\prod_{j=1}^{n+1}(t-\mu_j),$$

$$\boldsymbol{\Psi}_2(t)=\prod_{i=1}^{n}(t-\lambda_i).$$

因为 $\boldsymbol{\Psi}_1(t)$ 和 $\boldsymbol{\Psi}_2(t)$ 分别为 $n+1$ 和 n 阶多项式, 故存在数 c 及不超过 $n-1$ 阶的多项式 $\boldsymbol{\Psi}_3(t)$, 使得

$$\boldsymbol{\Psi}_1(t)=\boldsymbol{\Psi}_2(t)(t-c)+\boldsymbol{\Psi}_3(t). \tag{4.3.5}$$

(证明见蒋尔雄等 (1978), P.22). 直接计算表明: $c=\sum_{j=1}^{n+1}\mu_j-\sum_{i=1}^{n}\lambda_i=a$. 因为 $\boldsymbol{\Psi}_2(\lambda_i)=0, i=1,\cdots,n$, 从 (4.3.5) 得

$$\psi_1(\lambda_i)=\psi_3(\lambda_i),\quad i=1,\cdots,n$$

于是我们知道了 $\boldsymbol{\Psi}_3(t)$ 在 n 个点 $\lambda_1,\cdots,\lambda_n$ 处的值: $\boldsymbol{\Psi}_1(\lambda_i),\cdots,\boldsymbol{\Psi}_1(\lambda_n)$. 假设 $\lambda_1,\cdots,\lambda_n$ 互不相同, 应用 Lagrange 插值公式可得到

$$\boldsymbol{\Psi}_3(t)=\sum_{i=1}^{n}\boldsymbol{\Psi}_1(\lambda_i)\frac{\boldsymbol{\Psi}_2(t)}{\boldsymbol{\Psi}_2'(\lambda_1)(t-\lambda_i)}.$$

结合 (4.3.5), 有

$$\frac{\boldsymbol{\Psi}_1(t)}{\boldsymbol{\Psi}_2(t)} = (t-a) + \frac{\boldsymbol{\Psi}_3(t)}{\boldsymbol{\Psi}_2(t)} = (t-a) - \sum_{i=1}^{n} \frac{-\boldsymbol{\Psi}_1(\lambda_i)}{\boldsymbol{\Psi}_2'(\lambda_i)(t-\lambda_i)}.$$

结合 $\boldsymbol{\Psi}_1(\mu_j)=0, j=1,\cdots,n+1$, 则

$$(\mu_j - a) - \sum_{i=1}^{n} \frac{-\boldsymbol{\Psi}_1(\lambda_i)}{\boldsymbol{\Psi}_2'(\lambda_i)(\mu_j-\lambda_i)} = 0, \quad j=1,\cdots,n+1. \tag{4.3.6}$$

比较 (4.3.4) 和 (4.3.6) 知, 如果我们取

$$b_i^2 = -\frac{\boldsymbol{\Psi}_1(\lambda_i)}{\boldsymbol{\Psi}_2'(\lambda_i)}, \quad i=1,\cdots,n,$$

则 $\varphi_B(\mu_j)=0$, 即 $\mu_1,\cdots,\mu_{n+1}$ 为 B 的特征值, 现在剩下的问题是证明

$$\frac{\boldsymbol{\Psi}_1(\lambda_i)}{\boldsymbol{\Psi}_2'(\lambda_i)} \leqslant 0, \quad i=1,\cdots,n. \tag{4.3.7}$$

事实上, 根据假设 (4.3.2)

$$\boldsymbol{\Psi}_1(\lambda_i) = (-1)^i \sum_{j=1}^{n+1} |\lambda_i - \mu_j|,$$
$$\boldsymbol{\Psi}_2'(\lambda_i) = (-1)^{i-1} \sum_{\substack{j=1\\1\neq i}}^{n} |\lambda_i - \lambda_j|,$$

这就推出了 (4.3.7).

最后我们讨论 $\lambda_1,\cdots,\lambda_n$ 中, 至少有两个相等的情况. 假设 $\lambda_1=\cdots=\lambda_k > \lambda_{k+1} \geqslant \cdots$, 这里 $k \geqslant 2$. 根据假设 (4.3.2), 有 $\mu_2=\cdots=\mu_k=\lambda_1$. 此时 $\boldsymbol{\Psi}_1(t)$ 有一个因子 $(t-\mu_1)(t-\lambda_1)^{k-1}$, 而 $\boldsymbol{\Psi}_2(t)$ 有一个因子 $(t-\lambda_1)^k$, 且 λ_1 作为 $\boldsymbol{\Psi}_2(t)$ 的零点, 其重数为 k. 于是我们可以用

$$g_1(t) = \frac{\boldsymbol{\Psi}_1(t)}{(t-\lambda_1)^{k-1}},$$
$$g_2(t) = \frac{\boldsymbol{\Psi}_2(t)}{(t-\lambda_1)^{k-1}},$$

和

$$g_3(t) = \frac{\boldsymbol{\Psi}_3(t)}{(t-\lambda_1)^{k-1}}$$

分别代替原来的 $\boldsymbol{\Psi}_1(t)$, $\boldsymbol{\Psi}_2(t)$ 和 $\boldsymbol{\Psi}_3(t)$. 此时, λ_1 只是 $\boldsymbol{\Psi}_2(t)$ 之单根, 因此可以重复前面的讨论, 定理证毕.

在前面的讨论中, 我们是把矩阵 B 看成矩阵 A 镶边之后生成的镶边矩阵. 但是我们还可以从相反的角度来看这个问题, 把矩阵 A 看成从矩阵 B 剔除第 $n+1$ 行和列之后得到的主子阵, 于是, 关系 (4.3.2) 表征了 $n+1$ 阶 Hermite 阵与它的 n 阶主子阵的特征值之间的关系. 这里要说明, $n+1$ 阶 Hermite 阵 B 有 $n+1$ 个 n 阶主子阵, 虽然前面说的是剔除第 $n+1$ 行和列后的主子阵, 但 (4.3.2) 对任一 n 阶主子阵都成立. 这一断言的正确性容易从定理 4.3.1 的证明过程看出来.

进一步, 若我们考虑的不只是剔除某一行和列, 而是同时剔除 k 行和 k 列后的 $n-k$ 阶主子阵, 这是 Sturm 定理要回答的问题, 我们将在 §4.5 讨论.

§4.4 矩阵和的特征值

在应用上往往有这样的情况, 我们感兴趣的是方阵 A, 但实际上 A 不能被精确观测到, 我们只能观测到 $A+B$, 这里 B 是一个扰动矩阵或称误差矩阵, 在这种情况下, 我们就需要研究 $A+B$ 的特征值与 A 和 B 特征值之间的关系, 这方面结果的一个应用领域是统计诊断. 对此课题感兴趣的读者可参阅 Wang Songgui 和 Liski(1992).

定理 4.4.1 (Weyl) 设 A 和 B 皆为 $n\times n$ Hermite 阵. 则

$$\lambda_i(A)+\lambda_n(B)\leqslant\lambda_i(A+B)\leqslant\lambda_i(A)+\lambda_1(B),\quad i=1,\cdots,n. \tag{4.4.1}$$

证明 因为

$$\frac{x^*Ax}{x^*x}+\min_{x\neq0}\frac{x^*Bx}{x^*x}\leqslant\frac{x^*(A+B)x}{x^*x}\leqslant\frac{x^*Ax}{x^*x}+\max_{x\neq0}\frac{x^*Bx}{x^*x},$$

应用 Rayleigh-Ritz 定理 (定理 4.1.1), 我们有

$$\frac{x^*Ax}{x^*x}+\lambda_n(B)\leqslant\frac{x^*(A+B)x}{x^*x}\leqslant\frac{x^*Ax}{x^*x}+\lambda_1(B).$$

再应用 Courant-Fischer 定理 (定理 4.2.1), 即得 (4.4.1). 证毕.

推论 4.4.1 (单调性) 设 A 和 B 为两个 n 阶 Hermite 阵.

(1) 若 $A\geqslant B$, 则 $\lambda_i(A)\geqslant\lambda_i(B), i=1,\cdots,n$;

(2) 若 $A>B$, 则 $\lambda_i(A)>\lambda_i(B), i=1,\cdots,n$.

证明 (1) 因 $A\geqslant B$, 故 $C=A-B\geqslant0$. 对 $A=B+C$ 应用上面的定理, 得

$$\lambda_i(A)=\lambda_i(B+C)\geqslant\lambda_i(B)+\lambda_n(C)\geqslant\lambda_i(B),\quad i=1,\cdots,n.$$

最后一个不等式是因 $\lambda_n(C)\geqslant0$. 得证.

(2) 在 (1) 的证明中, 注意到对 $A>B$ 的情形, 有 $\lambda_n(C)>0$. 命题得证.

注 1　需要指出, 这两个事实的逆命题并不成立. 例如, 对情形 (1), 一个反例为

$$A=\begin{pmatrix}3&0\\0&2\end{pmatrix},\qquad B=\begin{pmatrix}1&0\\0&3\end{pmatrix}.$$

对情形 (2), 一个反例是

$$A=\begin{pmatrix}3&0\\0&2\end{pmatrix},\qquad B=\begin{pmatrix}1&0\\0&2\end{pmatrix}.$$

定理 4.4.2 (Weyl)　设 A 和 B 为 n 阶 Hermite 阵, 则

(1) $\lambda_i(A+B)\leqslant\lambda_j(A)+\lambda_k(B), i\geqslant j+k-1$;

(2) $\lambda_j(A)+\lambda_k(B)\leqslant\lambda_{j+k-n}(A+B), j+k\geqslant n$.

证明　(1) 设 $U=(u_1,\cdots,u_n)$ 和 $V=(v_1,\cdots,v_n)$ 为两个酉阵, 它们的列向量分别为 A 和 B 的特征值 $\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$ 和 $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$ 的特征向量. 应用定理 4.1.2, 有

$$\begin{aligned}\lambda_j(A)+\lambda_k(B)&=\max_{\substack{x\neq 0\\ x\perp u_t\\ t=1,\cdots,j-1}}\frac{x^*Ax}{x^*x}+\max_{\substack{x\neq 0\\ x\perp v_t\\ t=1,\cdots,k-1}}\frac{x^*Bx}{x^*x}\\ &\geqslant\max_{\substack{x\neq 0\\ x\perp u_t,t=1,\cdots,j-1\\ x\perp v_t,t=1,\cdots,k-1}}\frac{x^*(A+B)x}{x^*x}\geqslant\lambda_{j+k-1}(A+B)\\ &\geqslant\lambda_i(A+B),\end{aligned}$$

其中, 在倒数第二个不等号处应用了 Courant-Fischer 不等式 (4.2.1).

(2) 类似地, 我们有

$$\begin{aligned}\lambda_j(A)+\lambda_k(B)&=\min_{\substack{x\neq 0\\ x\perp u_t\\ t=j+1,\cdots,n}}\frac{x^*Ax}{x^*x}+\min_{\substack{x\neq 0\\ x\perp v_t\\ t=k+1,\cdots,n}}\frac{x^*Bx}{x^*x}\\ &\leqslant\min_{\substack{x\neq 0\\ x\perp u_t,t=j+1,\cdots,n\\ x\perp v_t,t=k+1,\cdots,n}}\frac{x^*(A+B)x}{x^*x}\leqslant\lambda_{j+k-n}(A+B).\end{aligned}$$

在最后一步我们应用了 Courant-Fischer 不等式 (4.2.2). 证毕.

为了证明定理 4.4.1 的一个推广结果, 我们先证如下引理.

引理 4.4.1　设 A 为 $n\times n$ Hermite 阵, $S_{j_1},\cdots,S_{j_k}$ 为 C^n 的任意 k 个子空间, 满足

$$S_{j_1}\subset\cdots\subset S_{j_k},\quad \dim S_{j_i}=j_i,\quad i=1,\cdots,k.$$

记

$$M=\{X_{n\times k}:X=(x_1,\cdots,x_k),X^*X=I_k,x_i\in S_{j_i},i=1,\cdots,k\},$$

则

$$\max_{\substack{x_i\in S_{j_i}\\ i=1,\cdots,k}}\min_{X\in M}\mathrm{tr}X^*AX=\sum_{i=1}^k\lambda_{j_i}(A). \tag{4.4.2}$$

证明 不失一般性, 我们可以假设 $A\geqslant 0$. 事实上对任意的 Hermite 阵 A, 存在 $d>0$, 使得

$$\begin{aligned}&A+dI_n>0,\\&\sum_{i=1}^k\lambda_{j_i}(A+dI_n)=\sum_{i=1}^k[\lambda_{j_i}(A)+d]=\sum_{i=1}^k\lambda_{j_i}(A)+kd.\end{aligned}$$

另一方面, 对任意 $X\in M$,

$$\mathrm{tr}X^*(A+dI_n)X=\mathrm{tr}X^*AX+kd.$$

可见, 我们只需对 $A\geqslant 0$ 证明 (4.4.2).

证明分如下两步.

(1) 先证 (4.4.2) 的左边 $\geqslant$ 右边.

记 $\varphi_1,\cdots,\varphi_n$ 为 A 的对应于特征值 $\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$ 的标准正交化特征向是, 定义

$$S_{j_i}=\mathscr{M}(\varphi_1,\cdots,\varphi_{j_i}),\quad i=1,\cdots,k.$$

因为 $j_1\leqslant\cdots\leqslant j_k\leqslant n$, 显然 $S_{j_1}\subset S_{j_2}\subset\cdots\subset S_{j_k},\dim(S_{j_i})=j_i,i=1,\cdots,k$. 记 $\boldsymbol{\Phi}_{(j_i)}=(\varphi_1,\cdots,\varphi_{j_i})$. 对任给的 $x_i\in S_{j_i},i=1,\cdots,k,\|x_i\|=1,x_i^*x_j=0,i\neq j$, 存在向量 t_i, 使得 $x_i=\boldsymbol{\Phi}_{(j_i)}t_i$, 于是

$$\begin{aligned}x_i^*Ax_i&=t_i^*\boldsymbol{\Phi}_{(j_i)}^*\boldsymbol{\Phi}\Lambda\boldsymbol{\Phi}^*\boldsymbol{\Phi}_{(j_i)}t_i\\&=t_i^*\mathrm{diag}(\lambda_1,\cdots,\lambda_{j_i})t_i\\&\geqslant\lambda_{j_i},\end{aligned}$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, 且当 $t_i=(0,\cdots,0,1)'$ 时, 等号成立. 因而

$$\sum_{i=1}^k\lambda_{j_i}\geqslant\min_{X\in M}\sum_{i=1}^k x_i^*Ax_i\leqslant\max_{\substack{S_{j_i}\\ i=1,\cdots,k}}\min_{U\in M}\mathrm{tr}U^*AU. \tag{4.4.3}$$

(2) 证明 (4.4.2) 的左边 $\leqslant$ 右边.

很明显, 我们需证明, 对任给的 k 个子空间

$$S_{j_1}\subset S_{j_2}\subset\cdots\subset S_{j_k},\quad \dim S_{j_i}=j_i,\quad i=1,\cdots,k,$$

必存在 $X=(x_1,\cdots,x_k)\in M$, 使得

$$\sum_{i=1}^{k}\lambda_{j_i}(A)\geqslant\sum_{i=1}^{k}x_i^*Ax_i\geqslant\min_{U\in M}\operatorname{tr}U^*AU. \tag{4.4.4}$$

我们可以把 A 看作 C^n 上的一个线性变换. 当一个子空间 $L\subset C^n$, 若对任意 $x\in L, Ax\in L$ 成立, 则限制在子空间 L 上, A 仍然看作一个线性变换. 记为 $A|_L$, $A|_L$ 的特征值与特征向量也是 A 的特征值和特征向量, 且

$$\lambda_i(A|_L)\leqslant\lambda_i(A). \tag{4.4.5}$$

关于这一事实的严格证明, 见孙继广 (1987, p. 124. 引理 3.3).

注意到 A 一定是包含 S_{j_k} 的某一子空间 S^m 上的变换, 这里 $\dim S^m=m$, 这是因为我们总可以选 $S^m=C^n$. 由 (4.4.5) 知, 为证 (4.4.4), 我们只需证明, 能够找到 $x_1,\cdots,x_k$, 使 $X=(x_1,\cdots,x_k)\in M$, 且

$$\sum_{i=1}^{k}\lambda_{j_i}(A|_{S^m})\geqslant\sum_{i=1}^{k}x_i^*Ax_i. \tag{4.4.6}$$

下面对 S^m 的维数 m 用归纳法.

(a) 设 $m=1$, 此时 $j_k=1, k=1$, 由 Rayleigh-Ritz 定理, 有

$$\lambda_1(A|_{S^1})=\max_{\substack{x\in S^1\\ \|x\|=1}}x^*Ax\geqslant x_1^*Ax_1,$$

其中 $x_1\in S^1=S_1, \|x_1\|=1$. 这就对 $k=1$ 证明了 (4.4.6).

(b) 设维数 $\leqslant m-1$ 时, (4.4.6) 成立. 我们证明维数 $=m$ 时, (4.4.6) 也成立. 因 $k\leqslant j_k\leqslant m$, 故这时仍需分两种情形.

1°. 若 $k=m$, 这时 $j_k=k$, 因而所有 $j_i=i, i=1,\cdots,k$. 取 $x_1\in S_1$, 且 $\|x_1\|=1$. 将其扩充为 S_2 的标准正交基 $\{x_1,x_2\}$, 再扩充为 S_3 的基 $\{x_1,x_2,x_3\}$. 一直这样做下去, 可以得到 $x_i\in S_i(i=1,\cdots,k)$, 且 $x_1,\cdots,x_k$ 为 S_k 的标准正交基. 记 $X=(x_1,\cdots,x_k)$, 有 $X^*X=I_k$, 利用定理 6.6.1, 有

$$\sum_{i=1}^{k}\lambda_i(A|_{S^m})=\max_{\substack{U^*U=I_k\\ u_i\in S^m}}\operatorname{tr}U^*AU\geqslant\sum_{i=1}^{k}x_i^*Ax_i.$$

此即 (4.4.6).

2°. 若 $k<m$, 这时又要分 $j_k<m$ 和 $j_k=m$ 两种情形来讨论.

(甲): $k<m, j_k<m$.

对这种情形, 有维数为 $m-1$ 的子空间 S^{m-1}, 满足

$$S_{j_k} \subset S^{m-1} \subset S^m,$$

取 $B=PAP$, 这里 P 为向子空间 S^{m-1} 的正交投影阵. 不难验证, B 为 S^{m-1} 上的线性变换, 由归纳假设

$$\sum_{i=1}^{k} \lambda_{j_i}(B|_{S^{m-1}}) \geqslant \sum_{i=1}^{k} x_i^* B x_i, \tag{4.4.7}$$

这里 $x_i \in S_{j_i}, \|x_i\|=1, x_i^* x_j = 0, i \neq j$. 利用 (4.4.5) 及推论 4.6.2 得

$$\begin{aligned}\lambda_{j_i}(B|_{S^{m-1}}) &= \lambda_{j_i}(PAP|_{S^{m-1}}) = \lambda_{j_i}(PA|_{S^{m-1}}) \\ &\leqslant \lambda_{j_i}(PA|_{S^m}) \leqslant \lambda_{j_i}(A|_{S^m}), \quad i=1,\cdots,k.\end{aligned} \tag{4.4.8}$$

因为所有 $x_i \in S^{m-1}$, 则

$$x_i^* B x_i = x_i^* PAP x_i = x_i^* A x_i, \quad i=1,\cdots,k.$$

所以, 综合 (4.4.7) 和 (4.4.8), 我们有

$$\sum_{i=1}^{k} \lambda_{j_i}(A|_{S^m}) \geqslant \sum_{i=1}^{k} \lambda_{j_i}(B|_{S^{m-1}}) \geqslant \sum_{i=1}^{k} x_i^* B x_i = \sum_{i=1}^{k} x_i^* A x_i.$$

于是, 对这种情况, 我们证明了 (4.4.6).

(乙): $k<m$, 且 $j_k=m$.

对于这种情形, $j_1 \leqslant j_2 \leqslant \cdots \leqslant j_k$ 必不是从 1 开始的彼此相邻的自然数. 故存在自然数介于某个 j_{l-1} 和 j_l 之间, 设 N 为具有此性质的自然数中最大者, 即 N 满足

$$\begin{cases} j_{l-1} < N < j_l, \quad \text{约定 } j_0 = 0, \\ (j_l, \cdots, j_k) = (N+1, \cdots, m), \end{cases}$$

此时可取 S^{m-1} 满足

$$S_{j_{l-1}} \subset S^{m-1} \subset S^m. \tag{4.4.9}$$

另外, 记 $\boldsymbol{\Psi}_{N+1}, \cdots, \Psi_m$ 为 $A|_{s^m}$ 的第 $N+1, \cdots, m$ 个特征值对应的特征向量. 由 (4.4.9) 我们有

$$S_{j_{l-1}} \subset (S^{m-1} \cap S_{j_l}) \subset \cdots \subset (S^{m-1} \cap S_{j_k}).$$

应用定理 1.1.1, 得

$$\begin{aligned}\dim(S^{m-1} \cap S_{j_i}) &= \dim S^{m-1} + \dim S_{j_i} - \dim(S^{m-1} + S_{j_i}) \\ &= m-1+j_i-\dim(S^{m-1}+S_{j_i}) \\ &\geqslant m-1+j_i-\dim(S^m) \\ &= j_i - 1, \qquad i=l,\cdots,k.\end{aligned}$$

故我们可以取 $S_N, S_{N+1}, \cdots, S_{m-1}$ 满足

$$S_{j_1} \subset \cdots \subset S_{j_{l-1}} \subset S_N \subset S_{N+1} \subset \cdots \subset S_{m-1},$$

这里, 当 $l=1$ 时, 约定 $S_{j_0} = \{0\}$.

仍与前面情形一样, 定义 B 与 P. 视 B 为 S^{m-1} 上的线性变换. 由归纳假设, 存在

$$x_i \in S_{j_i}, \quad i = 1, \cdots, l-1,$$

$$x_t \in S_t, \quad t = N, \cdots, m-1,$$

使得

$$\sum_{i=1}^{l-1} x_i^* B x_i + \sum_{i=1}^{m-1} x_i^* B x_i \leqslant \sum_{i=1}^{l-1} \lambda_{j_i}(B|_{S^{m-1}}) + \sum_{t=N}^{m-1} \lambda_t(B|_{S^{m-1}}). \tag{4.4.10}$$

同 (4.4.8) 的证明, 我们有

$$\lambda_{j_i}(B|_{S^{m-1}}) \leqslant \lambda_{j_i}(A|_{S^m}), \quad i = 1, \cdots, l-1. \tag{4.4.11}$$

而对 $t = N+1, \cdots, m$, 因 $\boldsymbol{\Psi}_t \in S^{m-1}, t = N+1, \cdots, m$. 故

$$B\boldsymbol{\Psi}_t = PAP\boldsymbol{\Psi}_t = PA\boldsymbol{\Psi}_t = \lambda_t(A|_{S^m})\boldsymbol{\Psi}_t,$$

这就是说, $\lambda_t(A|_{S^m})$ 是 B 作为 S^{m-1} 上变换的特征值, 注意到, $\lambda_N(B|_{S^{m-1}}), \cdots, \lambda_{m-1}(B|_{S^{m-1}})$ 是 $B|_{S^{m-1}}$ 上的最小的 $m-N$ 个特征值, 故必有

$$\sum_{t=N}^{m-1} \lambda_t(B|_{S^{m-1}}) \leqslant \sum_{t=N+1}^{m} \lambda_t(A|_{S^m}). \tag{4.4.12}$$

综合 (4.4.10), (4.4.11) 和 (4.4.12). 我们证明了

$$\sum_{i=1}^{l-1} x_i^* A x_i + \sum_{t=N}^{m-1} x_t^* A x_t \leqslant \sum_{i=1}^{k} \lambda_{j_i}(A|_{S^m}).$$

这就最终完成了 (4.4.6) 的证明. 引理证毕.

现在我们证明一个关于两个 Hermite 阵之和的特征值的一般性定理.

定理 4.4.3 设 A 和 B 为两个 $n \times n$ Hermite 阵, $C = A + B$ 且 $\alpha_1 \geqslant \cdots \geqslant \alpha_n, \beta_1 \geqslant \cdots \geqslant \beta_n$ 和 $\gamma_1 \geqslant \cdots \geqslant \gamma_n$ 分别为 A, B 和 C 的特征值. 则对任给的 $1 \leqslant j_1 \leqslant \cdots \leqslant j_k \leqslant n$, 有

$$\sum_{i=1}^{k} \alpha_{j_i} + \sum_{i=1}^{k} \beta_{n-k+i} \leqslant \sum_{i=1}^{k} \gamma_{j_i} \leqslant \sum_{i=1}^{k} \alpha_{j_i} + \sum_{i=1}^{k} \beta_i. \tag{4.4.13}$$

证明 根据引理 4.4.1, 存在 k 个子空间 $S_{j_1}\subset\cdots\subset S_{j_k}$, 使得

$$\sum_{i=1}^{k}\gamma_{j_i}=\min_{X\in M}\operatorname{tr}X^*CX, \tag{4.4.14}$$

这里 M 的定义同引理 4.4.1, 同时可以找到 $Y=(y_{j_1},\cdots,y_{j_k})$, 使得 $Y\in M$, 且

$$\min_{X\in M}\operatorname{tr}X^*AX=\operatorname{tr}Y^*AY.$$

但依引理 4.4.1, 有

$$\operatorname{tr}Y^*AY\leqslant\sum_{i=1}^{k}\alpha_{j_i}. \tag{4.4.15}$$

另一方面, 应用定理 6.6.1 知

$$\operatorname{tr}Y^*BY\leqslant\sum_{i=1}^{k}\beta_i. \tag{4.4.16}$$

综合 (4.4.14), (4.4.15) 和 (4.4.16), 我们得到

$$\sum_{i=1}^{k}\gamma_{j_i}\leqslant\operatorname{tr}Y^*CY=\operatorname{tr}Y^*AY+\operatorname{tr}Y^*BY\leqslant\sum_{i=1}^{k}\alpha_{j_i}+\sum_{i=1}^{k}\beta_i.$$

此即 (4.4.13) 的右边不等式.

为证 (4.4.13) 的左边不等式, 令 $D=-B$ 并记 $\delta_1\geqslant\cdots\geqslant\delta_n$ 为 D 的特征值, 利用上面已证结果, 有

$$\sum_{i=1}^{k}\alpha_{j_i}\leqslant\sum_{i=1}^{k}\gamma_{j_i}+\sum_{i=1}^{k}\delta_i,$$

注意到 $\delta_i=-\beta_{n-i+1},i=1,\cdots,k$, 这就得到 (4.4.13) 左边不等式. 定理证毕.

关于两个 Hermite 阵之和的一个重要特殊情形是 $B=uu^*$, 这里 u 为 $n\times 1$ 复向量, 即

$$C=A\pm uu^*.$$

这时, $B=uu^*$ 是秩为 1 的 Hermite 阵, 故 C 称为 A 的秩 1 修正. 在统计诊断中, 一组数据的影响分析往往归结到这种情况, 不过在统计学中我们遇到的都是实矩阵, 于是 A 是实对称阵, u 为实向量, 下面的定理给出了 $C=A\pm uu^*$ 与 A 的特征值的关系.

定理 4.4.4 设 A 为 $n\times n$ Hermite 阵, u 为任一向量, 则

(1) $\lambda_{i+2}(A+uu^*)\leqslant\lambda_{i+1}(A)\leqslant\lambda_i(A\pm uu^*)$,

$$i=1,2,\cdots,n-2;$$

(2) $\lambda_{i+2}(A) \leqslant \lambda_{i+1}(A \pm uu^*) \leqslant \lambda_i(A), i = 1, 2, \cdots, n-2$.

证明　当 $1 \leqslant i \leqslant n-1$ 时, 应用 Courant-Fischer 定理, 有

$$\begin{aligned}\lambda_i(A \pm uu^*) &= \min_{\substack{B \\ n\times(i-1)}} \max_{\substack{x\neq 0 \\ B^*x=0}} \frac{x^*(A \pm uu^*)x}{x^*x} \\ &\geqslant \min_{\substack{B \\ n\times(i-1)}} \max_{\substack{x\neq 0 \\ B^*x=0 \\ u^*x=0}} \frac{x^*(A \pm uu^*)x}{x^*x} \\ &= \min_{\substack{B \\ n\times(i-1)}} \max_{\substack{x\neq 0 \\ B^*x=0 \\ u^*x=0}} \frac{x^*Ax}{x^*x} = \min_{\substack{C_1 \\ c_i=u}} \max_{\substack{x\neq 0 \\ C^*x=0}} \frac{x^*Ax}{x^*x} \\ &\geqslant \min_{\substack{C \\ n\times i}} \max_{\substack{x\neq 0 \\ C^*x=0}} \frac{x^*Ax}{x^*x} = \lambda_{i+1}(A). \end{aligned} \tag{4.4.17}$$

其中 $C = (C_1, c_i)$, C_1 为 $n \times (i-1)$ 矩阵, c_i 为 $n \times 1$ 向量.

另一方面, 对 $i = 2, \cdots, n$, 再次应用 Courant-Fischer 定理, 得

$$\begin{aligned}\lambda_i(A \pm uu^*) &= \max_{\substack{B \\ n\times(n-i)}} \min_{\substack{x\neq 0 \\ B^*x=0}} \frac{x^*(A \pm uu^*)x}{x^*x} \\ &\leqslant \max_{\substack{B \\ n\times(n-i)}} \min_{\substack{x\neq 0 \\ B^*x=0 \\ u^*x=0}} \frac{x^*(A \pm uu^*)x}{x^*x} \\ &= \max_{\substack{C_1 \\ c_{n-i+1}=u}} \min_{\substack{x\neq 0 \\ C^*x=0}} \frac{x^*Ax}{x^*x},\end{aligned}$$

其中 $\underset{n\times(n-i+1)}{C} = (\underset{n\times(n-i)}{C_1}, \underset{n\times 1}{c_{n-i+1}})$

$$\leqslant \max_{C} \min_{\substack{x\neq 0 \\ C^*x=0}} \frac{x^*Ax}{x^*x} = \lambda_{i-1}(A). \tag{4.4.18}$$

综合 (4.4.17) 和 (4.4.18), 定理证毕.

注 2　在 (4.4.17) 的证明中, 对 $i = 1$, B 为 $n \times n-1$ 矩阵. 此时, 我们作如下理解

$$\begin{aligned}\lambda_1(A \pm uu^*) &= \min_{\substack{B \\ n\times(n-1)}} \max_{\substack{x\neq 0 \\ B^*x=0}} \frac{x^*(A \pm uu^*)x}{x^*x} \\ &= \max_{x\neq 0} \frac{x^*(A \pm uu^*)x}{x^*x}.\end{aligned}$$

在 (4.4.18) 证明中, 对 $i = n$, 也作类似理解.

在上一定理中, 我们已知 Hermite 阵 $B = uu^*$, 它的秩为 1. 但如果我们知道 $r(B) \leqslant t$, 那么我们有如下定理, 这些事实在积分方程理论中具有一定应用.

定理 4.4.5　设 A 和 B 皆为 $n \times n$ Hermite 阵, 且 $r(B) \leqslant t$. 则

(1) $\lambda_{i+2t}(A + B) \leqslant \lambda_{i+t}(A) \leqslant \lambda_i(A + B)$,　$i = 1, \cdots, n-2t$;

(2) $\lambda_{i+2t}(A) \leqslant \lambda_{i+t}(A+B) \leqslant \lambda_i(A), \quad i=1,\cdots,n-2t.$

证明 若记 $\lambda_1(B) \geqslant \cdots \geqslant \lambda_n(B)$ 为 B 的特征值, $\boldsymbol{\Psi}_1,\cdots,\boldsymbol{\Psi}_n$ 为对应的标准正交化特征向量. 因已知 $r(B) \leqslant t$, 故 $\lambda_{t+1}(B)=\cdots=\lambda_n(B)=0$, 因而 B 有谱分解

$$B=\sum_{i=1}^{t}\lambda_i(B)\boldsymbol{\Psi}_i\boldsymbol{\Psi}_i^*.$$

定理剩余部分的证明类似于上一定理, 只是将极值约束条件 $u^*x=0$ 用 $\boldsymbol{\Psi}_i^*x=0, i=1,\cdots,t$ 来代替. 定理证毕.

定理 4.4.6 设 A 和 B 为 n 阶正定 Hermite 阵, $A \geqslant B > 0$, 则

$$\frac{\lambda_k(B)}{\lambda_j(A)} \leqslant \frac{\lambda_k(A+B)}{\lambda_j(A+B)}. \tag{4.4.19}$$

证明 由 Courant-Fischer 定理 (定理 4.2.1), 有

$$\frac{\lambda_k(B)}{\lambda_j(A)}=\min_{\substack{R_{k-1}\\R_{n-j}}}\max_{\substack{R_{k-1}^*x=0\\R_{n-j}^*y=0}}\frac{x'Bx}{y'Ay}, \tag{4.4.20}$$

这里 R_m 为任意 $n\times m$ 的矩阵, 向量 x 和 y 满足 $x^*x=1$, $y^*y=1$. 由 $A \geqslant B > 0$, 我们易证

$$\frac{x'Bx}{y'Ay} \leqslant \frac{x'(A+B)x}{y'(A+B)y}, \quad \forall\ x,y\in R_n, \tag{4.4.21}$$

只要注意到 (4.4.21) 的等价形式为

$$x'Bx\cdot y'By \leqslant x'Ax\cdot y'Ay.$$

结合 (4.4.19) 和 (4.4.21), 即可得证.

§4.5 Sturm 定理

设 A 为 $n\times n$ 方阵. 对 $1\leqslant t\leqslant n, A(i_1,\cdots,i_t)$ 表示 A 的 t 阶主子阵这样的主子阵共有 $\begin{pmatrix} n \\ t \end{pmatrix}$ 个, 类似于上节研究镶边矩阵的技巧, 我们将应用 Courant-Fischer 定理建立主子阵 $A(i_1,\cdots,i_t)$ 的特征值与 A 的特征值之间的不等式. 因为这些不等式只依赖于 $A(i_1,\cdots,i_t)$ 的阶数 t, 而与具体的那个主子阵无关, 于是下面我们用 A_t 表示 A 的任一 t 阶主子阵.

定理 4.5.1 (Sturm) 设 A 为 $n\times n$ Hermite 阵, A_t 为任一 t 阶主子阵, 则

$$\lambda_{n-t+i}(A) \leqslant \lambda_i(A_t) \leqslant \lambda_i(A), \quad i=1,\cdots,t.$$

证明　假设 $A_t = A(i_1, \cdots, i_t)$, 应用 Courant-Fischer 定理, 有

$$\lambda_{n-t+i}(A) = \max_{\substack{B \\ n\times(t-i)}} \min_{\substack{x\neq 0 \\ B^*x=0}} \frac{x^*Ax}{x^*x} \leqslant \max_{\substack{B \\ n\times(t-i)}} \min_{\substack{x\neq 0 \\ B^*x=0 \\ x_j=0 \\ j\neq i_1,\cdots,i_t}} \frac{x^*Ax}{x^*x},$$

记 $y = (x_{i_1}, \cdots, x_{i_t})'$, $B' = (b_1, \cdots, b_n)$, $B_1 = (b_{i_1}, \cdots, b_{i_t})$, 则

$$\text{上式} = \max_{\substack{B_1 \\ t\times(t-i)}} \min_{\substack{y\neq 0 \\ B_1^*y=0}} \frac{y^*A_ty}{y^*y} = \lambda_i(A), \quad i = 1, \cdots, t.$$

另一方面, 完全类似地, 有

$$\lambda_i(A) = \min_{\substack{B \\ n\times(i-1)}} \max_{\substack{x\neq 0 \\ B^*x=0}} \frac{x^*Ax}{x^*x} \geqslant \min_{\substack{B \\ n\times(i-1)}} \max_{\substack{x\neq 0 \\ B^*x=0 \\ x_j=0 \\ j\neq i_1,\cdots,i_t}} \frac{x^*Ax}{x^*x} = \min_{\substack{B_1 \\ t\times(i-1)}} \max_{\substack{y\neq 0 \\ B_1^*y=0}} \frac{y^*A_ty}{y^*y} = \lambda_i(A_t), \quad i = 1, \cdots, t.$$

定理证毕.

§4.6　矩阵乘积的特征值

作为上节 Sturm 定理的一个应用, 我们首先导出著名的 Poincaré 分离定理.

定理 4.6.1 (Poincaré)　设 A 为 $n \times n$ Hermite 阵, U 为 $n \times k$ 列正交阵, 即 $U^*U = I_k$. 则

$$\lambda_{n-k+i}(A) \leqslant \lambda_i(U^*AU) \leqslant \lambda_i(A), \quad i = 1, \cdots, k. \tag{4.6.1}$$

设 $\boldsymbol{\Phi} = (\varphi_1, \cdots, \varphi_n)$ 为酉阵, 其列向量依次为 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ 对应的特征向量. 记 $\boldsymbol{\Phi}_{(k)}$ 和 $\boldsymbol{\Phi}_{[k]}$ 分别为 $\boldsymbol{\Phi}$ 的前 k 列和后 k 列组成的矩阵. 则 (4.6.1) 的右边不等式中等号对 $i = 1, \cdots, k$ 同时成立当且仅当 $U = \boldsymbol{\Phi}_{(k)}D$, 而 (4.6.1) 的左边不等式中等号对 $i = 1, \cdots, k$ 同时成立当且仅当 $U = \boldsymbol{\Phi}_{[k]}D$, 其中 D 为 $k \times k$ 酉阵.

证明　将 U 扩充为酉阵 $W = (U \vdots V)$. 记

$$\widetilde{A} = W^*AW = \begin{pmatrix} U^*AU & U^*AV \\ V^*AU & V^*AV \end{pmatrix}.$$

则

$$\lambda_i(\widetilde{A}) = \lambda_i(A), \quad i = 1, \cdots, n.$$

注意到 U^*AU 是 $\widetilde{A}$ 的 k 阶主子阵, 应用 Sturm 定理 (定理 4.5.1) 得

$$\lambda_{n-k+i}(A)=\lambda_{n-k+i}(\widetilde{A})\leqslant\lambda_i(U^*AU)\leqslant\lambda_i(\widetilde{A})=\lambda_i(A),\quad i=1,\cdots,k.$$

(4.6.1) 得证.

当 $U=\boldsymbol{\Phi}_{(k)}D$ 时,

$$U^*AU=D^*\boldsymbol{\Phi}_{(k)}^*A\boldsymbol{\Phi}_{(k)}D=D^*\mathrm{diag}(\lambda_1(A),\cdots,\lambda_k(A))D.$$

于是 $\lambda_i(U^*AU)=\lambda_i(A),i=1,\cdots,k$. 另一方面, 若 $\lambda_i(U^*AU)=\lambda_i(A),i=1,\cdots,k$, 则存在 k 阶酉阵 C_1, 使得

$$U^*AU=C_1\mathrm{diag}(\lambda_1(A),\cdots,\lambda_k(A))C_1^*,$$

即

$$C_1^*U^*AUC_1=\mathrm{diag}(\lambda_1(A),\cdots,\lambda_k(A)).$$

结合 $(UC_1)^*UC_1=I_k$, 上式表明, UC_1 的 k 个列为对应于 A 的前 k 个特征值的标准正交化特征向量, 于是存在 k 阶酉阵 C_2, 使得

$$UC_1=\boldsymbol{\Phi}_{(k)}C_2.$$

记 $D=C_2C_1^*$. 得 $U=\boldsymbol{\Phi}_{(k)}D$, 类似地, 可证明另一个充要条件. 定理证毕.

注 1 不等式 (4.6.1) 是由 Poincaré (1890) 首先证明的, 之所以把定理 4.6.1 称为分离定理, 是因为在不等式 (4.6.1) 中, 特征值 $\lambda_{n-k+i}(A)$ 和 $\lambda_i(A)$ 被 $\lambda_i(U^*AU)$ 分离开的缘故.

Poincaré 分离定理有着广泛的应用. Courant(1922) 在研究受线性约束的机械振动系统的振动频率问题时, 第一次系统研究和应用了 Poincaré 定理. Durbin 和 Watson(1950) 给出了 Poincaré 定理的第一个统计应用. 他们应用 Poincaré 定理证明了下面的推论 4.6.2, 并导出了回归分析中序列相关的著名 Durbin-WatsonD 检验 (陈希孺, 王松桂 1987, p.112).

推论 4.6.1 设 A 为 $n\times n$ 半正定 Hermite 阵, 则对一切满足 $X^*X=I_k$ 的矩阵 X, 有

$$\prod_{i=1}^{k}\lambda_{n-k+i}(A)\leqslant\det X^*AX\leqslant\prod_{i=1}^{k}\lambda_i(A),\tag{4.6.2}$$

右边等号成立当且仅当 $X=\boldsymbol{\Phi}_{(k)}\mathrm{D}$, 而左边等号成立当且仅当 $X=\boldsymbol{\Phi}_{[k]}D$, 这里 $\boldsymbol{\Phi}_{(k)}$, $\boldsymbol{\Phi}_{[k]}$ 和 D 的定义同定理 4.6.1.

注 2 在推论 4.6.1 中, 条件 "A 是半正定" 是必要的. 例如, 对

$$A=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&-2\end{pmatrix},\qquad X=\begin{pmatrix}0&0\\1&0\\0&1\end{pmatrix},$$

(4.6.2) 式就不成立.

推论 4.6.2 (Durbin-Watson)　设 A 为 $n\times n$ 半正定 Hermite 阵, P 为正交投影阵, $r(P)=k$. 则

$$\lambda_{n-k+i}(A)\leqslant\lambda_i(PA)\leqslant\lambda_i(A),\quad i=1,\cdots,k.\tag{4.6.3}$$

右边不等式中等号对一切的 $i=1,\cdots,k$ 同时成立当且仅当 $P=P_{\mathscr{M}(\boldsymbol{\Phi}_{(k)})}$. 而左边不等式中等号对一切 $i=1,\cdots,k$ 同时成立当且仅当 $P=P_{\mathscr{M}(\boldsymbol{\Phi}_{[k]})}$.

证明　因为 P 为正交投影阵. 且 $r(P)=k$, 应用定理 1.8.3, P 可表为 $P=QQ^*$, 这里 Q 满足 $Q^*Q=I_k$, 由定理 4.6.1, 我们有

$$\lambda_{n-k+i}(A)\leqslant\lambda_i(Q^*AQ)\leqslant\lambda_i(A),\quad i=1,\cdots,k,\tag{4.6.4}$$

因 $A\geqslant 0$, 于是 $Q^*AQ\geqslant 0$. 再应用定理 1.2.7 知, Q^*AQ 与 PA 有相同的正特征值, 故

$$\lambda_i(Q^*AQ)=\lambda_i(PA),\quad i=1,\cdots,k,\tag{4.6.5}$$

综合 (4.6.4) 和 (4.6.5), (4.6.3) 得证.

由定理 4.6.1 知, (4.6.4) 右边等号同时成立当且仅当 $Q=\boldsymbol{\Phi}_{(k)}D$, 这里 D 为 k 阶酉阵, 于是 $P=\boldsymbol{\Phi}_{(k)}\boldsymbol{\Phi}_{(k)}^*=P_{\mathscr{M}(\boldsymbol{\Phi}_{(k)})}$. 类似地, 可证另一充要条件, 推论证毕.

定理 4.6.2　设 A, B 为两个 n 阶 Hermite 阵.

(1) 若 $B\geqslant 0$, 则

$$\lambda_n(B)\lambda_i(A^2)\leqslant\lambda_i(ABA)\leqslant\lambda_1(B)\lambda_i(A^2),\quad i=1,\cdots,n;\tag{4.6.6}$$

(2) 若 $\lambda_1(B)>0,\lambda_n(B)<0$, 则

$$\lambda_n(B)\lambda_{n-i+1}(A^2)\leqslant\lambda_i(ABA)\leqslant\lambda_1(B)\lambda_i(A^2),\quad i=1,\cdots,n;\tag{4.6.7}$$

(3) 若 $B<0$, 则

$$\lambda_n(B)\lambda_{n-i+1}(A^2)\leqslant\lambda_i(ABA)\leqslant\lambda_1(B)\lambda_{n-i+1}(A^2),\quad i=1,\cdots,n.\tag{4.6.8}$$

证明　(1) 因为 $\lambda_1(B)I_n\geqslant B\geqslant\lambda_n(B)I$, 故

$$\lambda_1(B)A^2=A(\lambda_1(B)I_n-B)A+ABA\geqslant ABA,$$

$$ABA=A(B-\lambda_n(B)I_n)A+\lambda_n(B)A^2\geqslant\lambda_n(B)A^2,$$

应用推论 4.4.1, 得

$$\lambda_i(\lambda_1(B)A^2)\geqslant\lambda_i(ABA)\geqslant\lambda_i(\lambda_n(B)A^2).\tag{4.6.9}$$

当 $B \geqslant 0$ 时, $\lambda_i(B) \geqslant 0$, 从上式立得 (4.6.6).

其余两条结论可从 (4.6.9) 及下面的事实推出: 当 $\alpha < 0$ 时, $\lambda_i(\alpha A^2) = \alpha\lambda_{n-i+1} \cdot (A^2)$. 证毕.

推论 4.6.3 设 A 和 B 为 $n \times n$ 半正定 Hermite 阵, 则

$$\lambda_n(A)\lambda_i(B) \leqslant \lambda_i(AB) \leqslant \lambda_1(A)\lambda_i(B), \quad i = 1, \cdots, n; \tag{4.6.10}$$

$$\lambda_i(A)\lambda_n(B) \leqslant \lambda_i(AB) \leqslant \lambda_i(A)\lambda_1(B), \quad i = 1, \cdots, n. \tag{4.6.11}$$

证明 利用 $\lambda_i(AB) = \lambda_i(A^{1/2}BA^{1/2}) = \lambda_i(B^{1/2}AB^{1/2})$, 从 (4.6.6) 便得到所要结论.

定理 4.6.3 设 A 和 B 为两个任意 n 阶复方阵, $\lambda(AB)$ 为 AB 的任一特征值, 则

$$\lambda_n(A^*A)\lambda_n(B^*B) \leqslant |\lambda(AB)|^2 \leqslant \lambda_1(A^*A)\lambda_1(B^*B).$$

证明 设 φ 为特征值 $\lambda(AB)$ 对应的单位特征向量, 则

$$AB\varphi = \lambda(AB)\varphi,$$
$$\varphi^*B^*A^* = \overline{\lambda(AB)}\varphi^*,$$

于是

$$\varphi^*B^*A^*AB\varphi = |\lambda(AB)|^2.$$

应用 Ragleigh-Ritz 定理, 得

$$\lambda_1(B^*A^*AB) = \max_{x^*x=1} x^*B^*A^*ABx \geqslant |\lambda(AB)|^2,$$

$$\lambda_n(B^*A^*AB) = \min_{x^*x=1} x^*B^*A^*ABx \leqslant |\lambda(AB)|^2,$$

即

$$\lambda_n(B^*A^*AB) \leqslant |\lambda(AB)|^2 \leqslant \lambda_1(B^*A^*AB). \tag{4.6.12}$$

应用推论 4.6.3, 我们有

$$\lambda_1(B^*A^*AB) = \lambda_1(A^*ABB^*) \leqslant \lambda_1(A^*A)\lambda_1(B^*B), \tag{4.6.13}$$

$$\lambda_n(B^*A^*AB) = \lambda_n(A^*ABB^*) \geqslant \lambda_n(A^*A)\lambda_n(B^*B). \tag{4.6.14}$$

综合 (4.6.12), (4.6.13) 和 (4.6.14), 定理证毕.

注 3 记 $\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$ 和 $\sigma_1(B) \geqslant \cdots \geqslant \sigma_n(B)$ 分别为 A 和 B 的奇异值, 则定理 4.6.3 的结论可以改写为

$$\sigma_n(A)\sigma_n(B) \leqslant |\lambda(AB)| \leqslant \sigma_1(A)\sigma_1(B).$$

利用定理 4.6.2, 我们可以证明下面的定理, 它是 Poincaré 定理的一种推广.

定理 4.6.4　设 A 为 $n\times n$ 半正定 Hermite 阵, U 为 $n\times p$ 矩阵, 满足

$$U^*U=\varDelta=\operatorname{diag}(\delta_1,\cdots,\delta_p),\quad \delta_i>0,\quad i=1,\cdots,p,$$

则

$$\lambda_n(A)\delta_i\leqslant\lambda_i(U^*AU)\leqslant\lambda_1(A)\delta_i,\quad i=1,\cdots,p.$$

证明　记 $W=U\varDelta^{-1/2}$, 则 $W^*W=I_p$, 注意到

$$\lambda_i(U^*AU)=\lambda_i(\varDelta^{1/2}W^*AW\varDelta^{1/2}),$$

应用定理 4.6.2(1), 则有

$$\lambda_p(W^*AW)\delta_i\leqslant\lambda_i(U^*AU)\leqslant\lambda_1(W^*AW)\delta_i,\quad i=1,\cdots,p,\tag{4.6.15}$$

对 W^*AW 应用 Poincaré 定理, 得

$$\begin{aligned}\lambda_1(W^*AW)&\leqslant\lambda_1(A),\\ \lambda_p(W^*AW)&\geqslant\lambda_n(A),\end{aligned}$$

代入 (4.6.15), 定理得证.

注 4　在定理 4.6.4 条件下, 类似地, 我们可以证明

$$\delta_p\lambda_{n-p+i}(A)\leqslant\lambda_i(U^*AU)\leqslant\delta_1\lambda_i(A).$$

建议读者完成这一证明.

定理 4.6.5　设 A 和 B 为两个 $n\times n$ 半正这 Hermite 阵, $\mathscr{M}(A)\subset\mathscr{M}(B),X$ 为 $n\times k$ 矩阵, $q=r(B),t=r(BX)$, 则

$$\lambda_{q-t+i}(B^-A)\leqslant\lambda_i((X^*BX)^-X^*AX)\leqslant\lambda_i(B^-A),\quad i=1,\cdots,t,\tag{4.6.16}$$

这里所有特征值都与所包含的广义逆选择无关.

证明　首先由 $A\geqslant0,B\geqslant0$ 和 $\mathscr{M}(A)\subset\mathscr{M}(B)$ 容易推知, 对一切 $i,\lambda_i(B^-A)$ 与 $\lambda_i((X^*BX)^-X^*AX)$ 与所含广义逆选择无关. 因 $B\geqslant0$, 且秩为 q, 依满秩分解定理 (定理 1.5.3), B 可表为

$$B=GG^*,$$

其中, G 为 $n\times q$ 矩阵, $r(G)=q$, 因 BB^+ 为向 $\mathscr{M}(B)$ 的正交投影阵, $\mathscr{M}(A)\subset\mathscr{M}(B)$, 故

$$\begin{aligned}A&=BB^+A,\\ A&=AB^+B.\end{aligned}$$

因而

$$A = BB^+AB^+B = GG^*(GG^*)^+A(GG^*)^+GG^*.$$

利用 $G^+ = G^*(GG^*)^+$(见定理 1.7.10), 上式可改写为

$$A = GG^+A(GG^+)^*. \tag{4.6.17}$$

记 $E = G^+A(G^+)^*$, 这是半正定 Hermite 阵, 于是

$$A = GEG^*, \tag{4.6.18}$$

从而得到

$$(X^*BX)^-X^*AX = (X^*GG^*X)^-X^*GEG^*X. \tag{4.6.19}$$

注意到 $t = r(BX) = r(G^*X)$. 设

$$G^*X = PDQ^*$$

为 G^*X 的奇异值分解, 其中 D 为 $t \times t$ 非奇异对角阵, P 和 Q 分别为 $q \times t$ 和 $k \times t$ 矩阵, $P^*P = I_t, Q^*Q = I_t$, 在 (4.6.19) 中, 取 $(X^*BX)^-$ 为 $(X^*BX)^+$, 则

$$(X^*BX)^+X^*AX = QD^{-1}P^*EPD^*.$$

利用定理 1.2.7 及 Poincaré 定理得

$$\lambda_i((X^*BX)^-X^*AX) = \lambda_i((X^*BX)^+X^*AX) = \lambda_i(P^*EP), \quad i = 1, \cdots, t, \tag{4.6.20}$$

$$\lambda_{q-t+i}(E) \leqslant \lambda_i(P^*EP) \leqslant \lambda_i(E), \quad i = 1, \cdots, t. \tag{4.6.21}$$

另一方面

$$\lambda_i(E) = \lambda_i(G^+A(G^+)^*) = \lambda_i((GG^*)^+A) = \lambda_i(B^+A), \quad i = 1, \cdots, t, \tag{4.6.22}$$

综合 (4.6.20), (4.6.21) 和 (4.6.22), 定理得证.

注 5 这个定理是由 Scott 和 Styan(1985) 证明的, 他们把它应用于抽样调查数据的统计分析理论.

下面的推论是该定理的一个直接结果.

推论 4.6.4 设 A 和 B 为两个 $n \times n$ Hermite 阵, $A \geqslant 0, B > 0, X$ 为 $n \times k$ 矩阵, $r(X) = k$. 则

(1) $\displaystyle\sum_{i=1}^{k} \lambda_{n-k+i}(B^{-1}A) \leqslant \mathrm{tr}(X^*AX(X^*BX)^{-1}) \leqslant \sum_{i=1}^{k} \lambda_i(B^{-1}A)$;

(2) $\prod_{i=1}^{k}\lambda_{n-k+i}(B^{-1}A)\leqslant \det X^*AX/\det X^*BX\leqslant \prod_{i=1}^{k}\lambda_i(B^{-1}A).$

在结束这一节的时候, 我们再叙述几个关于两个半正定 Hermite 阵乘积的特征值的结果.

定理 4.6.6 设 A 和 B 为两个 $n\times n$ 半正定 Hermite 阵, 则

$$\sum_{i=1}^{k}\lambda_i(A)\lambda_{n-i+1}(B)\leqslant \sum_{i=1}^{k}\lambda_i(AB)\leqslant \sum_{i=1}^{k}\lambda_i(A)\lambda_i(B),$$
$$k=1,\cdots,n.$$

更进一步, 还有

定理 4.6.7 设 A 和 B 为 $n\times n$ 半正定 Hermite 阵, 则

$$\sum_{t=1}^{k}\lambda_{i_t}(AB)\geqslant \sum_{t=1}^{k}\lambda_{i_t}(A)\lambda_{n-t+1}(B),\quad 1\leqslant i_1<\cdots<i_k\leqslant n.$$

这两个定理的证明将在 §8.6 给出 (见定理 8.6.3 和定理 8.6.4).

定理 4.6.8 设 A 和 B 为 $n\times n$ 半正定 Hermite 阵, 则

$$\prod_{t=1}^{k}\lambda_{i_t}(AB)\leqslant \prod_{t=1}^{k}\lambda_{i_t}(A)\lambda_t(B),\quad 1\leqslant i_1<\cdots<i_k\leqslant n.$$

当 $k=n$ 时, 等号成立.

证明见 Lidskii(1950).

利用此定理, 我们可以很容易地证明如下事实.

定理 4.6.9 设 A 和 B 为 $n\times n$ 半正定 Hermite 阵, 则

$$\prod_{t=1}^{k}\lambda_t(AB)\geqslant \prod_{t=1}^{k}\lambda_{i_t}(A)\lambda_{n-i_t+1}(B),\quad 1\leqslant i_1<\cdots<i_k\leqslant n, \tag{4.6.23}$$

当 $k=n$ 时, 等号成立.

证明 先假设 A 和 B 为正定阵, 应用定理 4.6.8, 得

$$\begin{aligned}\prod_{t=1}^{k}\lambda_{i_t}(A)=\prod_{t=1}^{k}\lambda_{i_t}(B^{-1/2}AB^{1/2})&=\prod_{t=1}^{k}\lambda_{i_t}(B^{-1}B^{1/2}AB^{1/2})\\&\leqslant \prod_{t=1}^{k}\lambda_{i_t}(B^{-1})\lambda_t(B^{1/2}AB^{1/2})\\&=\prod_{t=1}^{k}\lambda_{i_t}(B^{-1})\lambda_t(AB)\\&=\prod_{t=1}^{k}\lambda^{-1}_{n-i_t+1}(B)\prod_{t=1}^{k}\lambda_t(AB).\end{aligned}$$

于是对 A 和 B 皆正定的情形, 我们证明了 (4.6.23).

若 A 和 B 为半正定阵, 则对任意 $\varepsilon > 0, A+\varepsilon I$ 和 $B+\varepsilon I$ 都是正定阵, 对它们利用已证事实可得到一个相应的不等式, 命 $\varepsilon \to 0$, 注意到矩阵的特征值是它的元素的连续函数, 便得到 (4.6.23). 证毕.

§4.7 特征值的界

本节我们将证明两个定理, 它们给出了一般方阵的特征值及其实部和虚部的界.

在 §3.7 我们已经指出, 任意 n 阶复方阵 A 可表为 $A = H + S$, 其中

$$H = \frac{1}{2}(A + A^*), \quad S = \frac{1}{2}(A - A^*)$$

分别称为 A 的 Hermite 部分和斜 Hermite 部分. 如果 A 本身就是 Hermite 阵, 则 $S = 0$, 因为此时 A 的所有特征值都是实数, 故

$$\sum_{i=1}^{n} |I_m(\lambda_i(A))| \leqslant \|S\| = 0. \tag{4.7.1}$$

这里 $I_m(\cdot)$ 表示复数的虚部, $\|S\|$ 表示 S 的任一种范数. (4.7.1) 表明, Hermite 阵 A 的斜 Hermite 部分的范数给出了 A 的特征值的虚部的模的上界, 类似地, 若 A 本身就是斜 Hermite 阵, 则 $H = 0$. 因为此时 A 的所有非零特征值皆为纯虚数 (见本节注 2 或许以超 (1963, p. 430, 定理 5)), 故有

$$\sum_{i=1}^{n} |\mathrm{Re}(\lambda_i(A))| \leqslant \|H\| = 0, \tag{4.7.2}$$

这里 $\mathrm{Re}(\cdot)$ 表示复数的实部. 因而 (4.7.2) 表明, 斜 Hermite 阵 A 的 Hermite 部分的范数给出了 A 的特征值实部的绝对值的上界. 下面的Schur 定理是这些特殊情形的推广.

定理 4.7.1 (Schur) 设 A 为 $n \times n$ 复方阵, $\|A\|_F$ 表示 A 的欧氏范数, 即 $\|A\|_F = (\mathrm{tr}(A^*A))^{1/2}, \lambda_1, \cdots, \lambda_n$ 为 A 的特征值, 则

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \|A\|_F^2; \tag{4.7.3}$$

$$\sum_{i=1}^{n} |\mathrm{Re}(\lambda_i)|^2 \leqslant \|H\|_F^2; \tag{4.7.4}$$

$$\sum_{i=1}^{n} |I_m(\lambda_i)|^2 \leqslant \|S\|_F^2. \tag{4.7.5}$$

若 (4.7.3), (4.7.4) 和 (4.7.5) 中有一个等号成立, 则其余两个等号也成立, 且等号成立的条件是 A 为规范阵.

证明 根据 Schur 分解 (定理 1.5.7), 存在 $n \times n$ 酉阵 U, 使得

$$A = U \varDelta U^*, \tag{4.7.6}$$

这里 $\varDelta$ 是一个 $n \times n$ 上三角阵, 其对角元为 $\lambda_1, \cdots, \lambda_n$. 显然, $\|A\|_F = \|\varDelta\|_F$. 记 $\varDelta$ 的非对角元为 $\delta_{ij}, j > i$. 则

$$\|A\|_F^2 = \|\varDelta\|_F^2 = \sum_{i=1}^n |\lambda_i|^2 + \sum_{j>i} |\delta_{ij}|^2 \geqslant \sum_{i=1}^n |\lambda_i|^2,$$

这就证明了 (4.7.3).

根据 (4.7.6), 我们有

$$H = \frac{1}{2} U (\varDelta + \varDelta^*) U^*.$$

于是

$$\|H\|_F^2 = \sum_{i=1}^n \left| \frac{\lambda_i + \overline{\lambda}_i}{2} \right|^2 + \sum_{j \neq i} \left| \frac{\delta_{ij} + \overline{\delta}_{ji}}{2} \right|^2 \geqslant \sum_{i=1}^n |\mathrm{Re}(\lambda_i)|^2,$$

(4.7.4) 得证, 类似地, 可证明 (4.7.5).

从上面的证明过程可以看出, 在 (4.7.3), (4.7.4) 和 (4.7.5) 中, 任何一个等号成立当且仅当所有 $\delta_{ij} = 0$, 等价地, A 酉相似于对角阵. 根据定理 1.5.5, 这等价于 A 为规范阵, 定理证毕.

注 1 利用这个定理很容易证明推论 3.3.2, 即对任意 n 阶复方阵

$$|\det A| \leqslant M^n n^{n/2},$$

这里 $M = \max_{i,j}\{|a_{ij}|\}$.

事实上, 由算术平均与几何平均不等式及本定理, 立得

$$\begin{aligned} |\det A|^2 = \prod_{i=1}^n |\lambda_i|^2 &\leqslant \left(\frac{1}{n} \sum_{i=1}^n |\lambda_i|^2 \right)^n \leqslant \left(\frac{1}{n} \|A\|_F^2 \right)^n \\ &\leqslant \left(\frac{1}{n} n^2 M^2 \right)^n. \end{aligned}$$

命题得证.

推论 4.7.1 设 $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值, $\mu_1, \cdots, \mu_n$ 为 A^*A 的特征值, 也就是 A 的奇异值的平方, $\sigma_i^2(A)(i = 1, \cdots, n)$, 则

$$\sum_{i=1}^n |\lambda_i|^2 \leqslant \sum_{i=1}^n \mu_i = \sum_{i=1}^n \sigma_i^2(A), \tag{4.7.7}$$

等号成立当且仅当 A 为规范阵.

结论容易从 (4.7.3) 得出, 在 §8.4, 我们将给出 (4.7.7) 的另一种证明.

定理 4.7.2 (Hirsch) 设 $A=(a_{ij})$ 为 $n\times n$ 复方阵, 记

$$M_1=\max_{i,j}\{|a_{ij}|\},$$

$$M_2=\max_{i,j}\{|a_{ij}+\bar{a}_{ji}|/2\},$$

$$M_3=\max_{i,j}\{|a_{ij}-\bar{a}_{ji}|/2\},$$

其中 $\bar{a}_{ij}$ 表示 a_{ij} 的共轭. 则对 A 的任一特征值 λ, 有

$$|\lambda|\leqslant nM_1; \tag{4.7.8}$$

$$|\mathrm{Re}(\lambda)|\leqslant nM_2; \tag{4.7.9}$$

$$|\mathrm{Im}(\lambda)|\leqslant nM_3. \tag{4.7.10}$$

证明 从不等式 (4.7.3) 得

$$|\lambda|\leqslant\left(\sum_{i=1}^{n}|\lambda_i|^2\right)^{1/2}\leqslant\|A\|_F\leqslant M_1,$$

于是 (4.7.8) 得证. 为证 (4.7.9) 和 (4.7.10), 注意到

$$M_2=\max_{i,j}\{|h_{ij}|\},$$
$$M_3=\max_{i,j}\{|s_{ij}|\},$$

这里 h_{ij} 和 s_{ij} 由下式定义

$$H=(h_{ij})=\frac{1}{2}(A+A^*),$$

$$S=(s_{ij})=\frac{1}{2}(A-A^*).$$

利用与上面类似的方法及 (4.7.4) 和 (4.7.5), 便可证明所要结论.

注 2 从 (4.7.10) 知, $M_3=0$ 蕴含着 $\mathrm{I_m}(\lambda)=0$. 另一方面 $M_3=0$ 当且仅当 A 是 Hermite 阵. 于是从 (4.7.10) 可立即推出：Hermit 阵的所有特征值都是实数. 类似地从 (4.7.9) 可推出：斜 Hermite 阵的所有特征值都是纯虚数.

推论 4.7.2 (Bendixson) 设 $A=(a_{ij})$ 为 $n\times n$ 实方阵, λ 为 A 的任一特征值, 记

$$d=\max_{ij}\{|a_{ij}-a_{ji}|/2\}.$$

则

$$|\mathrm{Im}(\lambda)| \leqslant \sqrt{\frac{n(n-1)}{2}}d.$$

证明　记 $\lambda_1,\cdots,\lambda_n$ 为 A 的特征值. 由 (4.7.5) 得

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{i\neq j}\left|\frac{a_{ij}-a_{ji}}{2}\right|^2 \leqslant n(n-1)d^2.$$

因为 A 是实方阵, 所以它的任一特征值的共轭复数也一定是 A 的特征值. 于是上式左端和式中的每个非零项一定是出现两次, 因而

$$2|\mathrm{Im}(\lambda_i)|^2 \leqslant n(n-1)d^2, \quad i=1,\cdots,n.$$

定理得证.

注 3　如果复方阵 A 的特征值都是实数, 则我们可以通过它的迹给出它们的界, 详见引理 5.3.3.

§4.8　Gerŝgorin 圆盘

设 $A=(a_{ij})$ 为 $n\times n$ 阵, 将 A 分解为 $A=D+B$, 这里 $D=\mathrm{diag}(a_{11},\cdots,a_{nn})$, 它是由 A 的对角元组成的对角阵, B 是把 A 的对角元都换成零之后得到的方阵, 若记

$$A_\varepsilon = D+\varepsilon B, \tag{4.8.1}$$

则 $A_0=D, A_1=A$. 很明显, $A_0=D$ 的特征值就是 $a_{11},\cdots,a_{nn}$. 因为方阵的特征值是其元素的连续函数, 所以, 当 ε 很小时, 我们有理由认为 A_ε 的特征值将落在 $a_{11},\cdots,a_{nn}$ 的邻域内. 下面的 Gerŝgorin 圆盘定理证明了, 这些邻域是以 $a_{11},\cdots,a_{nn}$ 为中心的圆盘.

定义 4.8.1　设 $A=(a_{ij})$ 为 $n\times n$ 阵. 记

$$R_i(A)=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|, \quad i=1,\cdots,n,$$

记

$$G_i(A)=\{z\colon |z-a_{ii}|\leqslant R_i(A)\},$$

称 $G_i(A)$ 为方阵 A 的第 i 个 Gerŝgorin 圆盘.

定理 4.8.1 (Gerŝgorin)　设 $A=(a_{ij})$ 为 $n\times n$ 阵, λ 为其任一特征值. 则

(1) $\lambda\in\bigcup_{i=1}^{n}G_i(A)$;

(2) 若 A 的 n 个圆盘 $G_i(A)$ 中, 有 m 个互相连通, 且与其余 $n-m$ 个不连通. 则在此 m 个圆盘组成的连通域中, 恰有 A 的 m 个特征值.

证明 (1) 用反证法. 若对 A 的任一特征值 λ, $\lambda \bar{\in} \bigcup_{i=1}^{n} G_i(A)$, 则

$$|\lambda - a_{ii}| > R_i(A), \quad i = 1, \cdots, n.$$

对方阵 $\lambda I - A$ 应用定理 3.11.1, 得到 $|\det(\lambda I - A)| \neq 0$, 这与 λ 是 A 的特征值相矛盾. 得证.

(2) 不失一般性, 我们可以假设 A 的前 m 个圆盘 $G_i(A)(i = 1, \cdots, m)$ 构成一个连通域, 并与后面的 $n-m$ 个圆盘分离开. 设方阵 A_ε 由 (4.8.1) 所定义, 记 $G_i(A_\varepsilon)(i = 1, \cdots, n)$ 为 A_ε 的 Gerŝgorin 圆盘, 则由已证的 (1) 知, 对 A_ε 的任一特征值 $\lambda(A_\varepsilon)$,

$$\lambda(A_\varepsilon) \in \bigcup_{i=1}^{n} G_i(A_\varepsilon),$$

并且对 $0 \leqslant \varepsilon \leqslant 1$ 有

$$G_i(A_\varepsilon) \subseteq G_i(A), G_i(A_0) = G_i(D) = a_{ii}, \quad i = 1, \cdots, n, \tag{4.8.2}$$

$$G_i(A_1) = G_i(A), \quad i = 1, \cdots, n. \tag{4.8.3}$$

因为 $\lambda_i(A_\varepsilon)$ 连续地依赖于 $\varepsilon \in [0,1]$, 因此当 ε 从 0 变化到 1 时, 每个 $\lambda_i(\varepsilon)$ 表示了复平面上一条连续曲线, 其起点为 $\lambda_i(0) = a_{ii}$, 终点为 $\lambda_i(1) = \lambda_i(A), i = 1, \cdots, n$. 由 (4.8.2), 当 ε 在 [0,1] 上连续变化时, A_ε 的每个圆盘始终在 A 的对应圆盘内, 于是由假设条件可推知, A_ε 的前 m 个圆盘始终与其后面的 $n-m$ 个圆盘分离, 故从 (4.8.2) 和 (4.8.3) 我们可以断言, 以 $a_{11}, \cdots, a_{nn}$ 为起点的 m 条连续曲线 $\lambda_1(A_\varepsilon), \cdots, \lambda_m(A_\varepsilon)$ 全部落在 $\bigcup_{i=1}^{m} G_i(A)$ 之内. 特别, 它们的终点 $\lambda_1(A), \cdots, \lambda_m(A) \in \bigcup_{i=1}^{m} G_i(A)$. 类似地可证, $\bigcup_{i=1}^{m} G_i(A)$ 不包含其余 $n-m$ 条曲线 $\lambda_i(A_\varepsilon), i = m+1, \cdots, n$, 因而也就不包含 $\lambda_{m+1}(A), \cdots, \lambda_n(A)$, 这就完成了定理证明.

下面我们看一个简单例子.

例 4.8.1 对方阵

$$A = \begin{pmatrix} 1 & 1 & 0 \\ \frac{1}{4} & 2 & \frac{1}{4} \\ \frac{1}{4} & 0 & 3 \end{pmatrix},$$

容易验证, 它的三个 Gerŝgorin 圆盘为

$$G_1(A)\colon |z-1|\leqslant 1;$$

$$G_2(A)\colon |z-2|\leqslant \frac{1}{2};$$

$$G_3(A)\colon |z-3|\leqslant \frac{1}{4}.$$

易见前两个圆盘是连通的, 而与 $G_3(A)$ 不连通. 根据定理我们可以断言, 在 $G_1(A)\bigcup G_2(A)$ 中有 A 的两个特征值, 而在 $G_3(A)$ 中有一个特征值.

因为 A 与 A' 有相同的特征值, 因此对 A' 应用上面的定理, 我们得到如下推论.

推论 4.8.1 设 $A=(a_{ij})$ 为 $n\times n$ 阵, λ 为其任一特征值, 则

(1) $\lambda\in\bigcup\limits_{i=1}^{n}G_i(A')$, 其中

$$G_i(A')=\{z\colon |z-a_{ii}|\leqslant C_i(A)\},$$

$$G_i(A)=\sum_{\substack{i=1\\ i\neq j}}^{n}|a_{ij}|;$$

(2) 若 A' 的 m 个圆盘 $G_i(A')$ 互相连通, 而与其余 $n-m$ 个不连通, 则在此 m 个圆盘组成的连通域中, 恰有 A 的 m 个特征值.

因为对任一可逆阵 $P, P^{-1}AP$ 与 A 有相同的特征值, 因此我们也可以对 $P^{-1}AP$ 应用定理 4.8.1 和推论 4.8.1, 得到 A 的特征值所在区域的估计, 适当地选择 P, 可望得到较好的估计. 特别, 选 P 为对角阵 $D=\mathrm{diag}(d_1,\cdots,d_n), d_i>0, i=1,\cdots,n$ 时, $D^{-1}AD=\left(\dfrac{d_ja_{ij}}{d_i}\right)_{n\times n}$, 对 $D^{-1}AD$ 应用前面的定理与推论, 我们得到

推论 4.8.2 设 $A=(a_{ij})$ 为 $n\times n$ 阵, λ 为 A 的任一特征值. $D=\mathrm{diag}(d_1,\cdots,d_n)$, $d_i>0, i=1,\cdots,n$. 则

$$\begin{aligned}\lambda&\in\bigcup_{i=1}^{n}G_i(D^{-1}AD)\\&=\bigcup_{i=1}^{n}\left\{z\colon |z-a_{ii}|\leqslant\frac{1}{d_j}\sum_{\substack{j=1\\ j\neq i}}^{n}d_j|a_{ij}|\right\},\end{aligned}$$

同时也有

$$\begin{aligned}\lambda&\in\bigcup_{i=1}^{n}G_i((D^{-1}AD)')\\&=\bigcup_{i=1}^{n}\left\{z\colon |z-a_{ii}|\leqslant d_j\sum_{\substack{i=1\\ i\neq j}}^{n}\frac{1}{d_i}|a_{ij}|\right\}.\end{aligned}$$

§4.9 Wielandt 不等式

在 §1.9, 我们讨论了 Cauchy-Schwarz 不等式的如下推广形式

$$|x^*Ay|^2 \leqslant x^*Ax \cdot y^*Ay, \tag{4.9.1}$$

其中, A 为 $n\times n$ 半正定 Hermite 阵, x, y 为任意一对 n 维向量. 因为当 x 与 y 成比例时, 等号成立, 可见如果对 x 与 y 没有附加限制, Cauchy-Schwarz 不等式 (4.9.1) 是不能再改进了. 本节要讨论的 Wielandt 不等式, 要求 x 与 y 为一对正交向量, 它是对 Cauchy-Schwarz 不等式 (4.9.1) 的一种改进.

定理 4.9.1 (Wielandt) 设 A 为 $n\times n$ 正定 Hermite 阵, $\lambda_1 \geqslant \cdots \geqslant \lambda_n > 0$ 为 A 的特征值, 则对任意一对正交向量 x 和 y, 有

$$|x^*Ay|^2 \leqslant \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 x^*Ax \cdot y^*Ay, \tag{4.9.2}$$

且存在正交向量 x 与 y, 使 (4.9.2) 的等号成立.

证明 显然我们只需对 $\|x\|=1, \|y\|=1$ 的正交向量证明 (4.9.2). 设 x 与 y 为任一对标准正交向量, 定义

$$B = (x, y)^*A(x, y),$$

这是一个 2×2 正定 Hermite 阵, 记其特征值为 $\mu_1 \geqslant \mu_2 > 0$. 根据 Poincaré 定理 (见定理 4.6.1), 我们有

$$\lambda_1 \geqslant \mu_1 \geqslant \mu_2 \geqslant \lambda_n. \tag{4.9.3}$$

另一方面

$$\begin{aligned} 1-\frac{|x^*Ay|^2}{(x^*Ax)(y^*Ay)} &= 4\frac{x^*Axy^*Ay-|x^*Ay|^2}{(x^*Ax+y^*Ay)^2-(x^*Ax-y^*By)^2} \\ &= \frac{4\det B}{(\mathrm{tr}B)^2-(x^*Ax-y^*Ay)^2} = \frac{4\mu_1\mu_2}{(\mu_1+\mu_2)^2-(x^*Ax-y^*Ay)^2} \\ &\geqslant \frac{4\mu_1\mu_2}{(\mu_1+\mu_2)^2}, \end{aligned} \tag{4.9.4}$$

这里等号成立当且仅当 $x^*Ax = y^*Ay$, 且 x, y 为一对标准正交向量. (4.9.4) 可以改写为

$$\frac{|x^*Ay|^2}{x^*Ax\cdot y^*Ay} \leqslant 1-\frac{4\mu_1\mu_2}{(\mu_1+\mu_2)^2} = \left(\frac{\mu_1-\mu_2}{\mu_1+\mu_2}\right)^2 = \left(\frac{\frac{\mu_1}{\mu_2}-1}{\frac{\mu_1}{\mu_2}+1}\right)^2,$$

因为右端是 μ_1/μ_2 的单调增函数, 结合 (4.9.3), 得

$$\frac{|x^*Ay|^2}{x^*Ax\cdot y^*Ay}\leqslant\left(\frac{\dfrac{\lambda_1}{\lambda_n}-1}{\dfrac{\lambda_1}{\lambda_n}+1}\right)^2=\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2,$$

(4.9.2) 得证.

若记 φ_1 和 φ_n 分别为对应于 λ_1 和 λ_n 的 A 的标准正交化特征向量, 则容易验证, 当 $x=(\varphi_1+\varphi_n)/\sqrt{2}, y=(\varphi_1-\varphi_n)/\sqrt{2}$ 时, 等号成立. 定理证毕.

最后我们给出 Wielandt 不等式的几何解释. 因 $A>0$, 故可分解为 $A=B^*B$. 若用 $\cos(Bx,By)$ 表示量 Bx 与 By 夹角 (指锐角) 的余弦, 那么

$$\cos(Bx,By)=\frac{|(Bx,By)|}{\|Bx\|\|By\|}=\frac{|x^*Ay|}{\sqrt{x^*Ax\cdot y^*Ay}}.$$

于是 Cauchy-Schwarz 不等式 (4.9.1) 等价于

$$\cos(Bx,By)\leqslant 1. \tag{4.9.5}$$

为了给出 Wieldant 不等式相对应的几何解释, 我们定义方阵 A 的条件数

$$\kappa(A)=\frac{\lambda_1}{\lambda_n}. \tag{4.9.6}$$

再定义 $\theta=\theta(A), 0\leqslant\theta\leqslant\dfrac{\pi}{2}$, 使得 $\cot\dfrac{\theta}{2}=\kappa^2(A)$, 因为

$$\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2=\left(\frac{\kappa^2(A)-1}{\kappa^2(A)+1}\right)^2=\left(\frac{\cot^2\dfrac{\theta}{2}-1}{\cot^2\dfrac{\theta}{2}+1}\right)^2$$

$$=\left(\frac{\cos^2\dfrac{\theta}{2}-\sin^2\dfrac{\theta}{2}}{\cos^2\dfrac{\theta}{2}+\sin^2\dfrac{\theta}{2}}\right)^2=\cos^2\theta.$$

于是 Wielandt 不等式等价于

$$\cos(Bx,By)\leqslant\cos\theta. \tag{4.9.7}$$

这就是说, 对于任意两个正交向量 x 与 y, Bx 与 By 所夹锐角不超过 $\theta=\theta(A)$. 当条件数 $\kappa(A)$ 很大时, 我们称方阵 A 呈病态. 因此, 方阵 A 病态程度愈严重, $\kappa(A)$ 就愈大, θ 也就愈大, 此时 $\cos\theta$ 就愈接近于 1, Wieldant 不等式 (4.9.7) 与 Cauchy-Schwarz 不等式 (4.9.5) 也就愈相近. 这表明, 对良态方阵 A(即 $\kappa(A)$ 比较小), Wieldant 不等式较 Cauchy-Schwarz 不等式有较大改进.

§4.10 Kantorovich 不等式及其推广

设 A 为正定 Hermite 阵, 在 §1.9 我们证明了 Cauchy-Schwarz 不等式的推广形式

$$|x^*y|^2 \leqslant x^*Ax \cdot y^*A^{-1}y. \tag{4.10.1}$$

特别当 $x=y$ 时,

$$(x^*x)^2 \leqslant x^*Ax \cdot x^*A^{-1}x, \tag{4.10.2}$$

等价地, 对任意 $x \neq 0$,

$$1 \leqslant \frac{x^*Axx^*A^{-1}x}{(x^*x)^2}. \tag{4.10.3}$$

上式右端是两个 Rayleigh 商

$$\frac{x^*Ax}{x^*x} \text{ 与 } \frac{x^*A^{-1}x}{x^*x}$$

的乘积. Cauchy-Schwarz 不等式 (4.10.3) 给出了这个乘积的下界 1. 本节我们要建立它的上界, 这就是下面的 Kantorovich 不等式.

定理 4.10.1 (Kantorovich) 设 A 为 $n \times n$ 正定 Hermite 阵, λ_1 和 λ_n 分别为其最大和最小特征值, 则对任意非零向量 x,

$$\frac{x^*Axx^*A^{-1}x}{(x^*x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \tag{4.10.4}$$

当 $x=(\varphi_1+\varphi_n)/\sqrt{2}$ 时, 等号成立, 这里 φ_1 和 φ_n 分别为 λ_1 和 λ_n 对应的标准正交化特征向量.

证明 设 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值, $\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n)$. 则存在酉阵 U, 使 $A=U^*\Lambda U$. 记 $y=Ux$.

$$\xi_i = |y_i|^2 / \left(\sum_{i=1}^n |\overline{y}_i|^2\right)^{1/2}, \quad i=1,2,\cdots,n,$$

问题归结为对 $\xi_i \geqslant 0, \sum\limits_{i=1}^n \xi_i = 1$, 证明

$$\left(\sum_{i=1}^n \lambda_i\xi_i\right)\sum_{i=1}^n\left(\frac{\xi_i}{\lambda_i}\right) \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}. \tag{4.10.5}$$

用下式定义 u_i 和 $v_i (i=1,\cdots,n)$.

$$\begin{cases} \lambda_i = \lambda_1 u_i + \lambda_n v_i, \\ \dfrac{1}{\lambda_i} = \dfrac{u_i}{\lambda_1} + \dfrac{v_i}{\lambda_n}. \end{cases} \tag{4.10.6}$$

容易验证, $u_i \geqslant 0, v_i \geqslant 0, i = 1, \cdots, n$.
再由

$$
\begin{aligned}
1 &= \frac{1}{\lambda_i}\lambda_i = \left(\frac{u_i}{\lambda_1} + \frac{v_i}{\lambda_n}\right)(\lambda_1 u_i + \lambda_n v_i) \\
&= (u_i + v_i)^2 + \frac{v_i u_i (\lambda_1 - \lambda_n)^2}{\lambda_1 \lambda_n},
\end{aligned}
$$

可推得 $u_i + v_i \leqslant 1, i = 1, \cdots, n$.

记

$$
u = \sum_{i=1}^{n} \xi_i u_i,
$$

$$
v = \sum_{i=1}^{n} \xi_i v_i,
$$

则有

$$
u + v = \sum_{i=1}^{n} \xi_i (u_i + v_i) \leqslant \sum_{i=1}^{n} \xi_i = 1. \tag{4.10.7}
$$

于是

$$
\begin{aligned}
\left(\sum_{i=1}^{n} \lambda_i \xi_i\right)\left(\sum_{i=1}^{n} \frac{\xi_i}{\lambda_i}\right) &= (\lambda_1 u + \lambda_n v)\left(\frac{u}{\lambda_1} + \frac{v}{\lambda_n}\right) \\
&= (u+v)^2 + uv\frac{(\lambda_1 - \lambda_n)^2}{\lambda_1 \lambda_n} \\
&= (u+v)^2 \left[1 + \frac{4uv}{(u+v)^2}\frac{(\lambda_1 - \lambda_n)^2}{4\lambda_1 \lambda_n}\right] \\
&\leqslant 1 + \frac{(\lambda_1 - \lambda_n)^2}{4\lambda_1 \lambda_n} = \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}.
\end{aligned}
$$

于是 (4.10.5) 得证, 容易验证, 当 $x = (\varphi_1 + \varphi_n)/\sqrt{2}$ 时, 等号成立, 定理证毕.

综合 (4.10.3) 和 (4.10.4), 我们有

$$
1 \leqslant \frac{x^* A x x^* A^{-1} x}{(x^* x)^2} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}. \tag{4.10.8}
$$

另一方面, (4.10.4) 可以改写为

$$
x^* A x \cdot x^* A^{-1} x \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}(x^* x)^2. \tag{4.10.9}
$$

从这个意义上说, Kantorovich 不等式 (4.10.9) 是 Cauchy-Schwarz 不等式 (4.10.2) 的 “逆” 形式.

注 1　Kantorovich 不等式还有许多种证明. 作为上节的 Wielandt 不等式的一个应用, 下面扼要介绍另外一种证法.

令 $y = \|x\|^2(A^{-1}x) - (x^*A^{-1}x)x$, 它满足 $x^*y = 0$, 且 (4.10.10)

$$Ay = \|x\|^2x - (x^*A^{-1}x)Ax, \tag{4.10.11}$$

$$x^*Ay = \|x\|^4 - (x^*A^{-1}x)(x^*Ax), \tag{4.10.12}$$

$$y^*Ay = -(x^*A^{-1}x)(y^*Ax). \tag{4.10.13}$$

从 (4.10.13) 立即推得 $y^*Ax = x^*Ay \leqslant 0$. 将 Wieldant 不等式 (4.9.7) 改写为

$$|x^*Ay|^2 \leqslant \cos^2\theta x^*Ax \cdot y^*Ay. \tag{4.10.14}$$

将 (4.10.13) 代入上式, 得

$$|x^*Ay|^2 \leqslant \cos^2\theta x^*Axx^*A^{-1}x(-y^*Ax).$$

因 $x^*Ay \leqslant 0$, 于是

$$-x^*Ay \leqslant \cos^2\theta x^*Axx^*A^{-1}x,$$

在 (4.10.12) 中, 利用这个不等式, 得到

$$\|x\|^4 \geqslant (1-\cos^2\theta)x^*Axx^*A^{-1}x$$

$$= \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}x^*Axx^*A^{-1}x,$$

此即 (4.10.4).

下面的定理是 Kantorovich 不等式的一个简单推广.

定理 4.10.2 (Greub-Rheinboldt) 设 A 和 B 为两个正定 Hermite 阵, 且 $AB = BA$, 记 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 和 $\mu_1 \geqslant \cdots \geqslant \mu_n$ 分别为 A 和 B 的特征值. 则对任意非零向量 x, 有

$$\frac{x^*A^2x \cdot x^*B^2x}{(x^*ABx)^2} \leqslant \frac{(\lambda_1\mu_1 + \lambda_n\mu_n)^2}{4\lambda_1\lambda_n\mu_1\mu_n}. \tag{4.10.15}$$

证明 因为 $AB = BA$, 根据定理 1.4.5, 存在酉阵 U, 使得 $A = U\Lambda U^*, B = UMU^*$, 这里 $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n), M = \operatorname{diag}(\mu_{i_1}, \cdots, \mu_{i_n})$. 命 $z = (\Lambda M)^{1/2}U^*x, C = \Lambda M^{-1} = \operatorname{diag}\left(\frac{\lambda_1}{\mu_{i_1}}, \cdots, \frac{\lambda_n}{\mu_{i_n}}\right)$. 应用 Kantorovich 不等式得

$$\frac{x^*A^2xx^*B^2x}{(x^*ABx)^2} = \frac{z^*Czz^*C^{-1}x}{(z^*z)^2} \leqslant \frac{(\delta_1+\delta_n)^2}{4\delta_1\delta_n},$$

其中 $\delta_1 = \max\limits_k \left\{ \dfrac{\lambda_k}{\mu_{i_k}} \right\}, \delta_n = \min\limits_k \left\{ \dfrac{\lambda_k}{\mu_{i_k}} \right\}$. 记上式右端为 d, 则

$$d = \frac{(\delta_1 + \delta_n)^2}{4\delta_1\delta_n} = \frac{\left(1 + \dfrac{\delta_1}{\delta_n}\right)^2}{4\left(\dfrac{\delta_1}{\delta_n}\right)}.$$

注意, d 是 $\dfrac{\delta_1}{\delta_n}$ 的单调增函数, 若记 $\alpha_1 = \dfrac{\mu_1}{\lambda_n}, \alpha_n = \dfrac{\mu_n}{\lambda_1}$, 从 δ_1 和 δ_n 的定义, 知

$$\frac{\alpha_1}{\alpha_n} \geqslant \frac{\delta_1}{\delta_n}.$$

于是

$$d \leqslant \frac{\left(1 + \dfrac{\alpha_1}{\alpha_n}\right)^2}{4\dfrac{\alpha_1}{\alpha_n}} = \frac{(\lambda_1\mu_1 + \lambda_n\mu_n)^2}{4\lambda_1\lambda_n\mu_1\mu_n}.$$

定理证毕.

现在我们考虑 Kantorovich 不等式的进一步推广. 在 (4.10.4) 中, 若假设 $x^*x = 1$, 则 Kantorovich 不等式变为

$$x^*Axx^*A^{-1}x \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}. \tag{4.10.16}$$

进一步的推广是将 $n \times 1$ 向量 x 换成 $n \times p$ 矩阵 X, 然后取其行列式, 给出 $\det X^*AX \cdot \det X^*A^{-1}X$ 的上界. 关于这一点, Bloomfield 和 Watson(1975) 在研究线性模型参数估计效率问题时证明了下面的定理, §9.2 将给出它的统计应用 (见定理 9.2.1).

定理 4.10.3 (Bloomfield-Watson)　设 A 为 $n \times n$ 实对称正定阵, X 为 $n \times p$ 实矩阵满足 $X'X = I_p$. 记 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值. 又 $n \geqslant 2p$. 则

$$\det X'AX \cdot \det X'A^{-1}X \leqslant \prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}. \tag{4.10.17}$$

证明　我们应用 Largrange 乘子法来证明. 记 $\varDelta$ 为 $p \times p$ 上三角阵, 它的 $\dfrac{1}{2}p(p+1)$ 个非零元素对应于约束条件 $X'X - I_p = 0$, 定义

$$F(X, \varDelta) = \ln\det X'A^{-1}X + \ln\det X'AX - 2\mathrm{tr}(X'X\varDelta).$$

利用例 1.11.5 和例 1.11.8, 我们有

$$\frac{\partial \ln\det X'A^{-1}X}{\partial X} = 2A^{-1}X(X'A^{-1}X)^{-1},$$

$$\frac{\partial \ln\det X'AX}{\partial X} = 2AX(X'AX)^{-1},$$
$$\frac{\partial \mathrm{tr}(X'X\Delta)}{\partial X} = X(\Delta+\Delta'),$$

对 $F(X,\Delta)$ 关于 X 求导数, 并令其等于 0, 得

$$AX(X'AX)^{-1} + A^{-1}X(X'A^{-1}X)^{-1} - X(\Delta+\Delta') = 0. \tag{4.10.18}$$

用 X 左乘上式可推出

$$\Delta+\Delta' = 2I_p,$$

代入 (4.10.18), 我们有

$$AX(X'AX)^{-1} + A^{-1}X(X'A^{-1}X)^{-1} = 2X. \tag{4.10.19}$$

再用 $X'A$ 左乘上式

$$X'A^2X(X'AX)^{-1} = 2X'AX - (X'A^{-1}X)^{-1}. \tag{4.10.20}$$

因为上式右边两个矩阵是对称阵, 于是左边的矩阵 $X'A^2X$ 和 $(X'AX)^{-1}$ 是可交换的, 因而 $X'A^2X$ 和 $X'AX$ 也是可交换的, 应用定理 1.3.5, 存在正交阵使 $X'A^2X$ 和 $X'AX$ 同时对角化. 再由 (4.10.20) 可推知, $X'AX$ 与 $X'A^{-1}X$ 也可以用同一正交阵同时对角化, 因为将 X 右乘一正交阵之后, (4.10.17) 的左端保持不变, 故我们假设 $X'AX$ 和 $X'A^{-1}X$ 已经是对角阵了.

记 $X=(x_1,\cdots,x_p)$, 则

$$X'AX = \mathrm{diag}(x_1'Ax_1,\cdots,x_p'Ax_p),$$
$$X'A^{-1}X = \mathrm{diag}(x_1'A^{-1}x_1,\cdots,x_p'A^{-1}x_p).$$

再记 (4.10.17) 左端为 $M(X)$, 于是

$$M(X) = \prod_{i=1}^{p} x_i'Ax_ix_i'A^{-1}x_i. \tag{4.10.21}$$

同时, (4.10.19) 变形为

$$\frac{Ax_i}{x_i'Ax_i} + \frac{A^{-1}x_i}{x_i'A^{-1}x_i} = 2x_i, \quad i=1,\cdots,p. \tag{4.10.22}$$

将上式左乘 A, 得

$$\frac{A^2x_i}{x_i'Ax_i} + \frac{x_i}{x_i'A^{-1}x_i} - 2Ax_i = 0, \quad i=1,\cdots,p. \tag{4.10.23}$$

由此可以推知, 对每个 i, x_i 和 Ax_i 位于最多由 A 的两个特征向量张成的子空间. 事实上, 设 $\varphi_1, \cdots, \varphi_n$ 为 A 的对应于 $\lambda_1, \cdots, \lambda_n$ 的标准正交化特征向量. 则 x_i 可表为 $x_i = \sum_{j=1}^{n} \alpha_{ij}\varphi_j, i = 1, \cdots, p$. 于是

$$Ax_i = \sum_{j=1}^{n} \alpha_{ij}\lambda_j\varphi_j,$$

$$A^2x_i = \sum_{j=1}^{n} \alpha_{ij}\lambda_j^2\varphi_j.$$

代入 (4.10.23), 我们有

$$\begin{aligned}
&\frac{A^2x_i}{x_i'Ax_i} + \frac{x_i}{x_i'A^{-1}x_i} - 2Ax_i = 0, \quad i = 1, \cdots, p \\
\Longleftrightarrow &\sum_{j=1}^{n} \alpha_{ij}\left(\frac{\lambda_j^2}{x_i'Ax_i} + \frac{1}{x_i'A^{-1}x_i} - 2\lambda_j\right)\varphi_j = 0, \quad i = 1, \cdots, p \\
\Longleftrightarrow &\alpha_{ij}\left(\frac{\lambda_j^2}{x_i'Ax_i} + \frac{1}{x_i'A^{-1}x_i} - 2\lambda_j\right) = 0, \quad i = 1, \cdots, p, \quad j = 1, \cdots, n.
\end{aligned}$$

于是对每个固定的 i, 若 $\alpha_{ij} \neq 0$, 则特征值 λ_j 必为二次方程

$$\frac{u^2}{x_i'Ax_i} - 2u + \frac{1}{x_i'A^{-1}x_i} = 0$$

的根. 因为此方程最多只有两个根, 所以对固定的 i, 最多只有两个 $\alpha_{ij} \neq 0$. 这就证明了每个 x_i 位于 A 的至多两个特征向量张成的子空间. 记对应的特征值为 a_i 和 b_i, 根据二次方程根与系数关系, 可得

$$\begin{cases}
a_i + b_i = 2x_i'Ax_i, & i = 1, \cdots, p, \\
a_ib_i = \dfrac{x_i'Ax_i}{x_i'A^{-1}x_i}, & i = 1, \cdots, p.
\end{cases}$$

于是

$$x_i'Ax_ix_i'A^{-1}x_i = \frac{(a_i + b_i)^2}{4a_ib_i}, \quad i = 1, \cdots, p.$$

容易看到 $x_i'A^{-1}x_ix_i'Ax_i$ 的最大值不会出现在 x_i 和 Ax_i 只落在 A 的一个特征向量张成的子空间的情形. 因为我们假设了 $X'X = I_p$, $X'AX$ 和 $X'A^2X$ 为对角阵, 所以向量偶 $\{x_1, Ax_1\}, \cdots, \{x_p, Ax_p\}$ 张成的 p 个子空间互相正交, 因而数偶

$\{a_1, b_1\}, \cdots, \{a_p, b_p\}$ 都彼此不同 (重根按重数计), 于是剩下的问题是从 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 中挑选出 $(a_1, \cdots, a_p)$ 和 $(b_1, \cdots, b_p)$, 使

$$M(X) = \prod_{i=1}^{p} x_i' A x_i x_i' A^{-1} x_i = \prod_{i=1}^{p} \frac{(a_i + b_i)^2}{4a_i b_i} \tag{4.10.24}$$

达到最大值.

为了使 (4.10.24) 达到最大, 我们应首先选取 λ_1 和 λ_n 配成对, 构成因子 $(\lambda_1 + \lambda_n)^2/(4\lambda_1\lambda_n)$. 其次, 再取 λ_2 和 λ_{n-1}, 构成因子 $(\lambda_2 + \lambda_{n-1})^2/(4\lambda_2\lambda_{n-1})$. 类推下去, 得到 $M(X)$ 的最大值

$$\prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}.$$

定理证毕.

推论 4.10.1 设 A 为 $n \times n$ 实对称正定阵, X 为 $n \times p$ 矩阵, 其秩为 p, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值, $n \geqslant 2p$. 则

$$\frac{\det X'AX \det X'A^{-1}X}{(\det X'X)^2} \leqslant \prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}. \tag{4.10.25}$$

证明 对 X 作 Schmidt 三角化分解 (见定理 1.5.1): $X = \widetilde{X}R$, 这里 $\widetilde{X}$ 为 $n \times p$ 矩阵, 满足 $\widetilde{X}'\widetilde{X} = I_p$. R 为 $p \times p$ 可逆上三角阵. 于是

$$\frac{\det X'AX \det X'A^{-1}X}{(\det X'X)^2} = \det \widetilde{X}'A\widetilde{X} \cdot \det \widetilde{X}'A^{-1}\widetilde{X},$$

应用定理 4.10.3. 推论得证.

注 2 关于 Kantorovich 不等式, 文献中还存在着许多其他形式的推广. 例如, Khatri 和 Rao(1981) 证明了

(1) $$\frac{\det X'AY \det Y'A^{-1}X}{\det X'X \cdot \det Y'Y} \leqslant \prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}},$$

其中 X 和 Y 皆为 $n \times p$ 矩阵, 且 $r(X) = r(Y) = p$. $n > 2p, \lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为正定阵 A 的特征值;

(2) $$\frac{\det X'A^2X \det X'B^2X}{(\det X'ABX)^2} \leqslant \prod_{i=1}^{p} \frac{(\mu_i + \mu_{n-i+1})^2}{4\mu_i\mu_{n-i+1}},$$

其中 A 与 B 为正定实对称阵, 且 $AB = BA, \mu_1 \geqslant \cdots \geqslant \mu_n$ 为 AB^{-1} 的特征值.

关于 Kantorovich 不等式的其他一些矩阵形式的推广及其应用, 将在 §7.6 和 §9.3 中讨论.

第5章　条　件　数

§5.1　定　　义

设 A 为 $n\times n$ 可逆阵. $\|A\|$ 表示 A 的任意一种范数, 则 A 的条件数定义为

$$\kappa(A)=\|A\|\|A^{-1}\|. \tag{5.1.1}$$

鉴于矩阵 A 的范数 $\|A\|$ 有多种定义, 于是由上式定义的条件数也就有许多种. 但最常用的条件数是由谱范数 $\|A\|_2$ 导出的, 称为谱条件数. 在本章中, 若无特殊声明, 我们讨论的条件数都是指谱条件数.

若记 $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$ 和 $\sigma_1(B)\geqslant\cdots\geqslant\sigma_n(B)$ 分别为方阵 B 的特征值和奇异值, 则 (谱) 条件数

$$\kappa(A)=\|A\|_2\|A^{-1}\|_2=\left(\frac{\lambda_1(A^*A)}{\lambda_n(A^*A)}\right)^{1/2}=\frac{\sigma_1(A)}{\sigma_n(A)}, \tag{5.1.2}$$

即一个可逆阵的条件数等于它的最大奇异值与最小奇异值之商. 若 A 为 Hermite 阵, 则

$$\kappa(A)=\frac{\lambda_1(A)}{\lambda_n(A)}. \tag{5.1.3}$$

假若 A 不是方阵, 而是 $m\times n$ 矩阵, $r(A)=t\leqslant n$, 则 A 的条件数定义为

$$\kappa(A)=\frac{\sigma_1(A)}{\sigma_t(A)},$$

即最大奇异值与最小非零奇异值之商.

一个可逆矩阵 $A=(a_1,\cdots,a_n)$ 称为病态的, 如果它的列向量 $a_1,\cdots,a_n$ 之间存在着近似线性关系. 此时 $\sigma_n(A)=\lambda_n^{1/2}(A^*A)$ 就会很小, 相应地条件数 $\kappa(A)$ 变得很大. 因此, 条件数可以用来度量矩阵的病态程度. 例如, 二阶方阵

$$A=\begin{pmatrix}1 & \varepsilon\\ \varepsilon & 1\end{pmatrix},\quad 其中\varepsilon>0.$$

容易验证：$\sigma_1(A)=1+\varepsilon,\sigma_2(A)=1-\varepsilon$. 当 $\varepsilon\to1$ 时,

$$\kappa(A)=\frac{1+\varepsilon}{1-\varepsilon}\to+\infty.$$

对应地,A 趋于奇异阵, 也就是说, A 的病态程度愈来愈严重.

条件数作为矩阵病态程度的数字度量, 在线性方程组解的稳定性研究中起着重要作用. 下面的定理具体地说明了这一点.

定理 5.1.1 设线性方程组 $Ax=b$ 有唯一解 x. 记 δA 和 δb 分别为 A 与 b 的扰动, 且满足 $\| A^{-1} \|_2 \| \delta A \|_2<1$. 则扰动后的线性方程组 $(A+\delta A)(x+\delta x)=b+\delta b$ 也有唯一解. 且

$$\frac{\| \delta x \|}{\| x \|} \leqslant \frac{\kappa(A)}{1-\kappa(A)\dfrac{\| \delta A \|_2}{\| A \|_2}} \left(\frac{\| \delta A \|_2}{\| A \|_2} + \frac{\| \delta b \|}{\| b \|} \right). \tag{5.1.4}$$

证明 因为 $\| A^{-1} \|_2 \| \delta A \|_2< 1$, 根据定理 4.6.3 知, $A^{-1}\delta A$ 的所有特征值都小于 1, 于是 $I+A^{-1}\delta A$ 的所有特征值都不会为零, 故 $\det(I+A^{-1}\delta A) \neq 0$. 但

$$\det(A+\delta A) = \det A(I+A^{-1}\delta A) = \det A \cdot \det(I+A^{-1}\delta A) \neq 0,$$

这表明方程组 $(A+\delta A)(x+\delta x)=b+\delta b$ 有唯一解.

因为

$$\begin{aligned} \delta x &= (A+\delta A)^{-1}(b+\delta b)-x \\ &= (A+\delta A)^{-1}\delta b - x + A^{-1}(I+\delta A\cdot A^{-1})^{-1}b. \end{aligned} \tag{5.1.5}$$

应用恒等式 $(I+C)^{-1}=I-(I+C)^{-1}C$, 得

$$(I+\delta A\cdot A^{-1})^{-1} = I-(I+\delta A\cdot A^{-1})^{-1}\delta A\cdot A^{-1}.$$

代入 (5.1.5), 我们有

$$\begin{aligned} \delta x &= (A+\delta A)^{-1}\delta b - x + A^{-}b - A^{-1}(I+\delta A\cdot A^{-1})^{-1}\delta A\cdot x \\ &= (A+\delta A)^{-1}\delta b - A^{-1}(I+\delta A\cdot A^{-1})^{-1}\delta A\cdot x \\ &= [(A+\delta A)^{-1} - A^{-1} + A^{-1}]\delta b - A^{-1}(I+\delta A\cdot A^{-1})^{-1}\delta A\cdot x, \end{aligned}$$

于是

$$\begin{aligned} \| \delta x \| \leqslant & (\| (A+\delta A)^{-1}-A^{-1} \|_2 + \| A^{-1} \|_2) \| \delta b \| \\ & + \| A^{-1} \|_2 \| (I+\delta A\cdot A^{-1})^{-1} \|_2 \| \delta A \|_2 \| x \| . \end{aligned} \tag{5.1.6}$$

再利用

$$(A+\delta A)^{-1}-A^{-1} = -A^{-1}\delta A(A+\delta A)^{-1},$$

得到

$$\begin{aligned} \| (A+\delta A)^{-1}-A^{-1} \|_2 \leqslant & \ \| A^{-1} \|_2 \| \delta A \|_2 \| (A+\delta A)^{-1} \|_2 \\ = & \ \| A^{-1} \|_2 \| \delta A \|_2 \| (I+A^{-1}\delta A)^{-1}A^{-1} \|_2 \end{aligned}$$

$$\leqslant \frac{\| A^{-1} \|_2^2 \| \delta A \|_2}{1 - \| A^{-1} \|_2 \| \delta A \|_2}, \tag{5.1.7}$$

这里, 在最后一个不等式处利用了定理 1.6.1. 同理

$$\| (I + \delta A \cdot A^{-1})^{-1} \|_2 \leqslant \frac{1}{1 - \| A^{-1} \|_2 \| \delta A \|_2}. \tag{5.1.8}$$

将 (5.1.7) 和 (5.1.8) 代入 (5.1.6), 得到

$$\begin{aligned} \| \delta x \| &\leqslant \frac{\| A^{-1} \|_2}{1 - \| A^{-1} \|_2 \| \delta A \|_2} \| \delta b \| + \frac{\| A^{-1} \|_2 \| \delta A \|_2}{1 - \| A^{-1} \|_2 \| \delta A \|_2} \| x \| \\ &= \frac{\| A^{-1} \|_2}{1 - \| A^{-1} \|_2 \| \delta A \|_2} (\| \delta b \| + \| \delta A \|_2 \| x \|). \end{aligned}$$

最后, 我们有

$$\begin{aligned} \frac{\| \delta x \|}{\| x \|} &\leqslant \frac{\kappa(A)}{1 - \kappa(A) \dfrac{\| \delta A \|_2}{\| A \|_2}} \left(\frac{\| \delta A \|_2}{\| A \|_2} + \frac{\| \delta b \|}{\| A \|_2 \| x \|} \right) \\ &\leqslant \frac{\kappa(A)}{1 - \kappa(A) \dfrac{\| \delta A \|_2}{\| A \|_2}} \left(\frac{\| \delta A \|_2}{\| A \|_2} + \frac{\| \delta b \|}{\| b \|} \right). \end{aligned}$$

定理得证.

(5.1.4) 表明, 条件数 $\kappa(A)$ 反映了线性方程组 $Ax = b$ 的解对于 A 和 b 的扰动的稳定程度. 因为 (5.1.4) 的右端是 $\kappa(A)$ 的单调增函数, 所以 $\kappa(A)$ 愈大, 对 A 与 b 的同样大小的扰动, $Ax = b$ 的解的相对改变量可能愈大. 关于这一点, 下面的例子会给读者留下一个直观印象.

例 5.1.1 (Wilson) 设线性方程组 $Ax = b$ 的系数矩阵 A 和常数向量 b 分别为

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \qquad b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix},$$

它的解为 $(x_1, x_2, x_3, x_4)' = (1, 1, 1, 1)'$. 如果把 b 扰动为 $b + \delta b = (32.1, 22.9, 33.1, 30.9)'$, 它的解就变为

$$x + \delta x = (9.2, -12.6, 4.5, -1.1)'.$$

此例表明, 对常数项 b 的微小相对扰动

$$\frac{\| \delta b \|}{\| b \|} \approx \frac{1}{200},$$

线性方程组解的相对改变量为 $\dfrac{\| \delta x \|}{\| b \|} \approx 2000$, 这充分说明了, 这个线性方程组的解关于常数项的扰动非常敏感. 类似的结论对系数阵 A 的扰动也成立. 例如, 若 b 保持不变, 而 A 扰动为

$$A+\delta A=\begin{pmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{pmatrix},$$

它的解变为

$$x+\delta x=(-81,137,-34,22)'.$$

与原来的 $x=(1,1,1,1)'$ 的相对改变量就更大了. 我们再来看 A 的条件数. 因为 A 为对称正定阵, $\lambda_1(A)=30.2887$, $\lambda_4(A)=0.01015$, 它的条件数

$$\kappa(A)=\frac{\lambda_1(A)}{\lambda_4(A)} \approx 2984.$$

它是如此之大, 这就难怪 A 与 b 的微小扰动会引起解的巨大变化. 这个例子充分直观地显示了在线性方程组求解时, 系数阵的条件数是何等重要!

为了减少线性方程组解的误差, 在应用上我们应当尽量减少系数阵和常数项的误差. 但这里需指出, 误差是不可避免的. 其原因有二：第一, 在应用上, 数据往往是从实际观测得到或是计算机算出的中间结果, 前者会有观测误差, 而后者会有计算误差; 其二, 无论我们采用的计算机多么先进, 它的字长总是有限的, 因而在原始数据输入机器时, 舍入误差是不可避免的. 鉴于这个原因, 当解线性方程组时我们必须特别注意系数阵的条件数. 有一种考虑是通过相似变换降低矩阵的条件数, 这称为“平衡”或“预处理”, 这是一个尚待深入研究的问题, 感兴趣的读者可参阅 Shapiro(1982) 和 McCarthy 等 (1973) 以及其后所列的参考文献.

在数理统计学中, 回归系数的最小二乘估计就是一个线性方程组 (即所谓正则方程) 的解, 因此读者不难明白, 条件数在最小二乘估计的稳定性研究中有重要应用, 这些属于回归诊断的研究范围 (详见 Besly 等 (1980)).

§5.2 性质与基本不等式

从定义容易验证, 条件数具有下列性质：

1. $\kappa(A) \geqslant 1$;
2. $\kappa(A)=\kappa(A^{-1})=\kappa(cA)$, 对任意 $c>0$;

3. 对任意酉阵 Q, $\kappa(QAQ^*)=\kappa(A)$;

4. $\kappa(AA^*)=\kappa^2(A)\geqslant\kappa(A)$.

最后这个事实证明比较容易. 事实上

$$\kappa(AA^*)=\frac{\sigma_1(AA^*)}{\sigma_n(AA^*)}=\kappa^2(A)\geqslant\kappa(A).$$

这个性质对所有其他范数定义的条件数都成立, 但证明不像现在谱条件数的情形这么容易 (见 Marshall 和 Olkin(1965)).

定理 5.2.1 设 A 和 B 为 $n\times n$ Hermite 正定阵, 则

$$\kappa(A+B)\leqslant\max\{\kappa(A),\ \kappa(B)\},$$

特别

$$\kappa(A+I)\leqslant\kappa(A).$$

证明 不妨设 $\kappa(A)\leqslant\kappa(B)$, 于是问题归结为证明

$$\kappa(A+B)\leqslant\kappa(B). \tag{5.2.1}$$

根据 Weyl 定理 (见定理 4.4.1), 有

$$\lambda_1(A+B)\leqslant\lambda_1(A)+\lambda_1(B),$$

$$\lambda_n(A+B)\geqslant\lambda_n(A)+\lambda_n(B).$$

于是

$$\kappa(A+B)=\frac{\lambda_1(A+B)}{\lambda_n(A+B)}\leqslant\frac{\lambda_1(A)+\lambda_1(B)}{\lambda_n(A)+\lambda_n(B)}. \tag{5.2.2}$$

再利用熟知的不等式: 设 a, b, c 和 d 皆为正数, 则

$$\frac{a}{b}<\frac{c}{d}\Longrightarrow\frac{a+c}{b+d}<\frac{c}{d}.$$

(5.2.1) 得证. 定理证毕.

注 1 我们用 $A\geqslant B$ 表示 $A-B\geqslant 0$, 一种范数 $\|A\|$ 称为单调的, 如果由 $A\geqslant B\geqslant 0$ 能推出 $\|A\|\geqslant\|B\|$. Marshall 和 Olkin(1965) 证明了所有酉不变范数都是单调范数, 且对所有由单调范数定义的条件数, 定理 5.2.1 都成立.

定理 5.2.2 设 A 为 $n\times n$ Hermite 正定阵, 且分块为

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},\quad \text{其中, } A_{11} \text{ 为 } r\times r \text{ 阵,}$$

则

(1) $\kappa(A) \geqslant \kappa(A_{11})$;

(2) $\kappa(A) \geqslant \kappa(A_{22} - A_{21}A_{11}^{-1}A_{12})$.

证明 (1) 应用 Sturm 定理 (见定理 4.5.1), 有

$$\lambda_1(A_{11}) \leqslant \lambda_1(A),$$
$$\lambda_r(A_{11}) \geqslant \lambda_n(A).$$

由此两式立得 $\kappa(A) \geqslant \kappa(A_{11})$.

(2) 由分块矩阵的逆矩阵公式 (见定理 1.7.11) 知, $(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$ 为 A^{-1} 的 $n-r$ 阶主子式, 利用 (1) 的结论, 得

$$\begin{aligned}\kappa(A_{22} - A_{21}A_{11}^{-1}A_{12}) &= \kappa((A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}) \\ &= \kappa((A^{-1})_{22}) \leqslant \kappa(A^{-1}) = \kappa(A),\end{aligned}$$

这里 $(A^{-1})_{22}$ 表示 A^{-1} 的右下角 $(n-r)\times(n-r)$ 的子阵. 定理证毕.

接下来的两个定理给出了矩阵乘积的条件数之间的关系.

定理 5.2.3 设 A 和 B 皆为 $n \times n$ Hermite 正定阵, 则

$$\max\left\{\frac{\kappa(A)}{\kappa(B)}, \frac{\kappa(B)}{\kappa(A)}\right\} \leqslant \kappa(AB) \leqslant \kappa(A)\cdot\kappa(B).$$

证明 应用定理 4.6.2, 得

$$\lambda_n^2(A)\lambda_i^2(B) \leqslant \lambda_i(BA^2B) \leqslant \lambda_1^2(A)\lambda_i^2(B), \quad i = 1, \cdots, n,$$
$$\lambda_i^2(A)\lambda_n^2(B) \leqslant \lambda_i(BA^2B) \leqslant \lambda_i^2(A)\lambda_1^2(B), \quad i = 1, \cdots, n.$$

于是, 对 AB 的最大奇异值 $\delta_1(AB)$, 有

$$\lambda_n(B)\lambda_1(A) \leqslant \delta_1(AB) = \lambda_1^{1/2}(BA^2B) \leqslant \lambda_1(A)\lambda_1(B),$$
$$\lambda_1(B)\lambda_n(A) \leqslant \delta_1(AB).$$

同时, 对 AB 的最小奇异值 $\delta_n(AB)$, 有

$$\lambda_n(A)\lambda_n(B) \leqslant \delta_n(AB) = \lambda_n^{1/2}(BA^2B) \leqslant \lambda_1(A)\lambda_n(B),$$
$$\delta_n(AB) \leqslant \lambda_n(A)\lambda_1(B).$$

因而, 下面的三个不等式同时成立

$$\kappa(AB) \leqslant \kappa(A)\kappa(B),$$

$$\kappa(AB) \geqslant \frac{\kappa(A)}{\kappa(B)},$$

$$\kappa(AB) \geqslant \frac{\kappa(B)}{\kappa(A)}.$$

定理得证.

定理 5.2.4 设 A 为 $n\times n$ Hermite 正定阵. X 为 $n\times p$ 矩阵, $R(X)=p$. 则

$$\frac{\lambda_{n-p+1}(A)}{\lambda_p(A)\kappa(X^*X)}\leqslant \kappa(X^*AX)\leqslant \kappa(X^*X)\kappa(A).$$

证明 容易看到, 我们的结论可以从下面的不等式推出

$$\lambda_{n-p+i}(A)\lambda_p(X^*X)\leqslant \lambda_i(X^*AX)\leqslant \lambda_i(A)\lambda_1(X^*X),\quad i=1,\cdots,p. \tag{5.2.3}$$

以下我们证明 (5.2.3).

因为对 $i=1,\cdots,p$,

$$\begin{aligned}\lambda_i(X^*AX\cdot\lambda_1(X^*X)^{-1}) &= \lambda_i(A^{1/2}X\cdot\lambda_1(X^*X)^{-1}\cdot X^*A^{1/2})\\ &\geqslant \lambda_i(A^{1/2}X(X^*X)^{-1}X^*A^{1/2})=\lambda_i(X^*AX\cdot(X^*X)^{-1}).\end{aligned}$$

故

$$\lambda_i(X^*AX)\geqslant \lambda_i(X^*AX(X^*X)^{-1})\lambda_p(X^*X),\quad i=1,\cdots,p. \tag{5.2.4}$$

应用定理 4.6.5, 得

$$\lambda_i(X^*AX(X^*X)^{-1})\geqslant \lambda_{n-p+i}(A),\quad i=1,\cdots,p. \tag{5.2.5}$$

综合 (5.2.4) 和 (5.2.5), 便得到 (5.2.3) 左边不等式.

另一方面

$$\begin{aligned}\lambda_i(X^*AX) &= \lambda_i(A^{1/2}XX^*A^{1/2})\leqslant \lambda_i(A^{1/2}\cdot\lambda_1(X^*X)I\cdot A^{1/2})\\ &= \lambda_i(\lambda_1(X^*X)A)=\lambda_1(X^*X)\lambda_i(A),\quad i=1,\cdots,p.\end{aligned}$$

此即 (5.2.3) 的右边不等式. 定理证毕.

定理 5.2.5 设 A 为 $n\times n$ 可逆阵. 则对任意 $n\times n$ 奇异阵 B,

$$\kappa(A)\geqslant \frac{\sigma_1(A)}{\sigma_1(A-B)}.$$

其中 $\sigma_1(\cdot)$ 表示最大奇异值.

证明 因为

$$B=A-(A-B)=A[I-A^{-1}(A-B)],$$

所以 $I-A^{-1}(A-B)$ 为奇异阵, 故有

$$||A^{-1}(A-B)||_2\geqslant 1.$$

于是从谱范数的相容性, 得

$$||A^{-1}||_2||A-B||_2 \geqslant 1,$$

因而

$$\sigma_1(A-B)=||A-B||_2 \geqslant \frac{1}{||A^{-1}||_2}=\frac{1}{\sigma_1(A^{-1})}=\sigma_n(A).$$

最后我们得到

$$\kappa(A)=\frac{\sigma_1(A)}{\sigma_n(A)} \geqslant \frac{\sigma_1(A)}{\sigma_1(A-B)}.$$

定理得证.

§5.3 条件数的界

我们并不总是需要知道给定矩阵的条件数的准确值. 事实上, 有时候只要我们能够估计出它的一个取值范围也就可以了. 为了这样的目的, 本节将给出方阵条件数上、下界的一些结果, 它们仅涉及一些方阵迹的运算, 应用起来非常方便.

设 A 为 $n\times n$ 复方阵, 假设它的特征值都是实数, 递减地排列为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. 记

$$M=\frac{1}{n}\sum_{j=1}^{n}\lambda_j=\frac{1}{n}\mathrm{tr}A, \tag{5.3.1}$$

$$\begin{aligned} S^2&=\frac{1}{n}\sum_{j=1}^{n}(\lambda_j-M)^2=\frac{1}{n}\left[\sum_{j=1}^{n}\lambda_j^2-\left(\sum_{j=1}^{n}\lambda_j\right)^2\right] \\ &=\frac{1}{n}\left[\mathrm{tr}A^2-\frac{1}{n}(\mathrm{tr}A)^2\right]=\frac{1}{n}\mathrm{tr}A^2-M^2. \end{aligned} \tag{5.3.2}$$

易见, M 和 S^2 分别为 A 的特征值 $\lambda_1,\cdots,\lambda_n$ 的算术平均值和方差, 因而 S 就是标准差. 下面的一些定理是通过 M 和 S 来给出条件数 $\kappa(A)$ 的上、下界.

我们先证明三个引理.

引理 5.3.1 设 x 和 y 为两个 $n\times 1$ 实向量, $\mathbf{1}=(1,\cdots,1)'$, 即 n 个元素皆为 $\mathbf{1}$ 的 n 维向量. 记

$$\begin{aligned} M_y&=y'\mathbf{1}/n, \\ S_y^2&=y'Cy/n, \end{aligned} \tag{5.3.3}$$

其中 $C=I_n-\dfrac{1}{n}\mathbf{1}\mathbf{1}'$. 则

$$-S_y(nx'Cx)^{1/2} \leqslant x'y-M_yx'\mathbf{1}=x'Cy \leqslant S_y(nx'Cx)^{1/2}, \tag{5.3.4}$$

且左边 (右边) 等号成立当且仅当存在数 $a<0(a>0)$ 及 b, 使得 $y=ax+b\mathbf{1}$.

证明　应用 Cauchy-Schwarz 不等式, 得

$$(x'Cy)^2 \leqslant x'Cx \cdot y'Cy,$$

由此立得 (5.3.4). 注意到 C 为实对称幂等阵, 故上式等号成立当且仅当 $Cy = aCx \Longleftrightarrow C(y-ax)=0$, 这等价于存在数 b, 使得 $y-ax=b\mathbf{1}$, 这就是 $y=ax+b\mathbf{1}$. 证毕.

引理 5.3.2　对任意一串数 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, 有

(1) $\lambda_n \leqslant M - \dfrac{S}{(n-1)^{1/2}} \leqslant M + \dfrac{S}{(n-1)^{1/2}} \leqslant \lambda_1$;

(2) 左边等号成立 $\Longleftrightarrow \lambda_2 = \cdots = \lambda_n$;

(3) 右边等号成立 $\Longleftrightarrow \lambda_1 = \cdots = \lambda_{n-1}$;

(4) 中间等号成立 $\Longleftrightarrow \lambda_1 = \cdots = \lambda_n$.

证明　因为

$$\begin{aligned} n^2(M-\lambda_n)^2 &= \left[\sum_{j=1}^{n}(\lambda_j-\lambda_n)\right]^2 \\ &= \sum_{j=1}^{n}(\lambda_j-\lambda_n)^2 + \sum_{j\neq k}(\lambda_j-\lambda_n)(\lambda_k-\lambda_n) \\ &\geqslant \sum_{j=1}^{n}(\lambda_j-\lambda_n)^2 = \sum_{j=1}^{n}(\lambda_j - M + M - \lambda_n)^2 \\ &= n[S^2+(M-\lambda_n)^2]. \end{aligned} \tag{5.3.5}$$

于是

$$(n^2-n)(M-\lambda_n)^2 \geqslant nS^2.$$

注意到 $\lambda_n < M$, 由上式立得 (1) 的左边不等式.

用同样的方法展开 $n^2(\lambda_1-M)^2$, 即可证明 (1) 的右边不等式.

在 (5.3.5) 中, 等号成立当且仅当

$$\sum_{j\neq k}(\lambda_j-\lambda_n)(\lambda_k-\lambda_n)=0 \Longleftrightarrow \lambda_2=\lambda_3=\cdots=\lambda_n,$$

因此证明了 (2), 其余结论的证明或是类似的, 或很显然. 证毕.

引理 5.3.3　设 A 为 $n\times n$ 复方阵, 其特征值全为实数, 且排列为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, 则

$$M-(n-1)^{1/2}S \leqslant \lambda_j \leqslant M+(n-1)^{1/2}S, \quad j=1,\cdots,n. \tag{5.3.6}$$

特别对 λ_1 和 λ_n 有

$$M+\frac{S}{(n-1)^{1/2}}\leqslant\lambda_1\leqslant M+(n-1)^{1/2}S, \tag{5.3.7}$$

$$M-(n-1)^{1/2}S\leqslant\lambda_n\leqslant M-\frac{S}{(n-1)^{1/2}}, \tag{5.3.8}$$

且 (5.3.7) 的左边 (右边) 等号成立 $\Longleftrightarrow$(5.3.8) 左边 (右边) 等号成立 $\Longleftrightarrow n-1$ 个最大 (最小) 特征值相等.

证明 在引理 5.3.1 中, 取 $x=e_j=(0,\cdots,0,1,0,\cdots,0)'$, 这里 1 位于第 j 个坐标, $y=(\lambda_1,\cdots,\lambda_n)'$, 则 (5.3.4) 变为

$$-(n-1)^{1/2}S\leqslant\lambda_j-M\leqslant(n-1)^{1/2}S,\quad j=1,\cdots,n.$$

(5.3.6) 得证. (5.3.7) 的左边不等式和 (5.3.8) 的右边不等式可从引理 5.3.2(1) 得出. 其余结论从引理 5.3.2(2) 和 (3) 推得. 证毕.

下面的定理给出了 Hermite 阵的条件数的界.

定理 5.3.1 (1) 设 A 为 $n\times n$ Hermite 正定阵, 则

$$\kappa(A)\geqslant 1+\frac{2\dfrac{S}{(n-1)^{1/2}}}{M-\dfrac{S}{(n-1)^{1/2}}}, \tag{5.3.9}$$

当 $n>2$ 时, 等号成立 $\Longleftrightarrow\lambda_1=\cdots=\lambda_n$.

(2) 设 A 为 $n\times n$ Hermite 阵, $\mathrm{tr}A>0$, $(\mathrm{tr}A)^2>(n-1)\mathrm{tr}A^2$. 则 $A>0$, 且

$$1+\frac{2\dfrac{S}{(n-1)^{1/2}}}{M-\dfrac{S}{(n-1)^{1/2}}}\leqslant\kappa(A)\leqslant 1+\frac{2S(n-1)^{1/2}}{M-(n-1)^{1/2}S}, \tag{5.3.10}$$

当 $n>2$ 时, 等号成立 $\Longleftrightarrow A=cI_n, c>0$.

证明 (1) 当 $A>0$ 时, 由 (5.3.8) 右边不等式知

$$M-\frac{S}{(n-1)^{1/2}}>0.$$

由 (5.3.7) 左边和 (5.3.8) 右边不等式便可推出 (5.3.9), 等号成立的充要条件由引理 5.3.3 直接得到.

(2) 因为

$$M-(n-1)^{1/2}S>0\Longleftrightarrow\mathrm{tr}A>0 \text{ 和 } (\mathrm{tr}A)^2>(n-1)\mathrm{tr}A^2.$$

再由 $M-(n-1)^{1/2}S>0 \Longrightarrow \lambda_n>0 \Longrightarrow A>0$. 这就证明了第一个结论.

因为在定理条件下, 有 $A>0$, 故从已证的 (1) 可推出 (5.3.10) 的左边不等式. (5.3.10) 的右边不等式由 (5.3.7) 的右边和 (5.3.8) 的左边不等式导出, 再由引理 5.3.3 中等号成立的条件可得到此处等号成立的条件. 定理证毕.

第6章　迹

无论从定义或计算角度, 与矩阵的秩、行列式、特征值和条件数等概念相比, 本章即将要讨论的矩阵的迹都是最简单的. 然而, 它在许多领域, 如数值计算、逼近论以及统计估计等都有相当多的应用. 许多量的计算都会归结到矩阵迹的运算. 本章我们将讨论有关迹的一些重要不等式.

§6.1　迹的基本性质

定义 6.1.1　设 $A=(a_{ij})$ 为 $n\times n$ 方阵, 称它的对角线元素之和为 A 的迹, 记为 $\mathrm{tr}A$. 即

$$\mathrm{tr}A=\sum_{i=1}^{n}a_{ii}.$$

容易验证, 迹具有下列性质：

(1) 线性：$\mathrm{tr}(\alpha A+\beta B)=\alpha\mathrm{tr}A+\beta\mathrm{tr}B$, 这里 α 和 β 皆为数;

(2) $\mathrm{tr}A'=\mathrm{tr}A$, $\mathrm{tr}A^*=\overline{\mathrm{tr}A}$;

(3) 设 A 和 B 分别为 $n\times m$ 和 $m\times n$ 矩阵, 则 $\mathrm{tr}AB=\mathrm{tr}BA$;

(4) 对任意可逆阵 P, $\mathrm{tr}P^{-1}AP=\mathrm{tr}A$;

(5) 设 A 为 $n\times n$ 方阵, x 为 $n\times 1$ 向量, 则 $x^*Ax=\mathrm{tr}Axx^*$;

(6) 设 $\lambda_1,\cdots,\lambda_n$ 为 A 的特征值, 则

$$\mathrm{tr}A=\sum_{i=1}^{n}\lambda_i, \tag{6.1.1}$$

$$\mathrm{tr}A^k=\sum_{i=1}^{n}\lambda_i^k; \tag{6.1.2}$$

(7) $\mathrm{tr}A^*A=0\Longleftrightarrow A=0$;

(8) 设 $A\geqslant 0$, 则 $\mathrm{tr}A\geqslant 0$, 且等号成立 $\Longleftrightarrow A=0$;

(9) 设 $A\geqslant B$(即 $A-B\geqslant 0$). 则 $\mathrm{tr}A\geqslant\mathrm{tr}B$, 且等号成立 $\Longleftrightarrow A=B$;

(6) 的证明从 Jordan 标准形及 (4) 推出. 另外 (6.1.1) 也可以用特征多项式来证明, 详见推论 1.2.1; (8) 是 (9) 的特殊情形; (9) 由推论 4.4.1 直接得到; (10) 设 A 为 $n\times n$ 方阵. 若存在正整数 k, 使得 $A^k=0$, 则 $\mathrm{tr}A=0$. 利用 Jordan 标准形, 容易完成此证明.

§6.2 若干基本不等式

对于两个 $m\times n$ 复阵 A 和 B, 迹 $\mathrm{tr}A^*B$ 是 $m\times n$ 维酉空间上的内积, 也就是将它们按列依次排成的两个 mn 维列向量的内积. 利用 Cauchy-Schwarz 不等式立即得到.

定理 6.2.1 对任意两个 $m\times n$ 矩阵 A 和 B, 有

$$|\mathrm{tr}A^*B|^2\leqslant \mathrm{tr}A^*A\cdot \mathrm{tr}B^*B,$$

这里等号成立 $\Longleftrightarrow A=cB$, c 为一数, 特别当 A 和 B 为实对称阵或 Hermite 阵时

$$0\leqslant |\mathrm{tr}AB|\leqslant (\mathrm{tr}A^2)^{1/2}(\mathrm{tr}B^2)^{1/2}.$$

定理 6.2.2 设 A 和 B 为两个 n 阶 Hermite 阵, 且 $A\geqslant 0, B\geqslant 0$. 则

$$0\leqslant \mathrm{tr}AB\leqslant \lambda_1(B)\mathrm{tr}A\leqslant \mathrm{tr}A\cdot\mathrm{tr}B, \tag{6.2.1}$$

这里 $\lambda_1(B)$ 表示 B 的最大特征值.

证明 利用上节迹的性质 (3) 和 (8), 我们有

$$\mathrm{tr}AB=\mathrm{tr}A^{1/2}BA^{1/2}\geqslant 0.$$

另一方面, 因为 $\lambda_1(A)B\geqslant B^{1/2}AB^{1/2}$, 利用上节性质 (9), 得

$$\mathrm{tr}AB=\mathrm{tr}A^{1/2}BA^{1/2}\leqslant \mathrm{tr}\lambda_1(A)B=\lambda_1(A)\mathrm{tr}B\leqslant \mathrm{tr}A\mathrm{tr}B.$$

证毕.

推论 6.2.1 设 A 为 $n\times n$ Hermite 方阵, 且 $A>0$. 则

$$\mathrm{tr}A\cdot\mathrm{tr}A^{-1}\geqslant n.$$

推论 6.2.2 设 A 和 B 为同阶 Hermite 阵, $A>0$, $B\geqslant 0$. 则

$$\mathrm{tr}A^{-1}B\geqslant \frac{\mathrm{tr}B}{\lambda_1(A)}\geqslant \frac{\mathrm{tr}B}{\mathrm{tr}A}, \tag{6.2.2}$$

等价地

$$\mathrm{tr}B\leqslant \lambda_1(A)\mathrm{tr}(A^{-1}B)\leqslant \mathrm{tr}A\cdot\mathrm{tr}(A^{-1}B). \tag{6.2.3}$$

证明 事实上, 应用定理 6.2.2, 有

$$\mathrm{tr}B=\mathrm{tr}(A\cdot A^{-1}B)\leqslant \mathrm{tr}(A^{-1}B)\lambda_1(A)\leqslant \mathrm{tr}A\cdot\mathrm{tr}(A^{-1}B).$$

证毕.

推论 6.2.3 设 A 为 $n\times n$ 半正定 Hermite 阵, 则对任意的正整数 k, 有

$$\mathrm{tr}(A^k)\leqslant(\mathrm{tr}(A))^k.$$

证明 对 k 实施归纳法. 在定理 6.2.2 中, 令 $B=A$, 由 (6.2.1) 得

$$\mathrm{tr}(A^2)\leqslant(\mathrm{tr}(A))^2.$$

于是命题对 $k=2$ 成立. 假设对正整数 $k-1$ 有

$$\mathrm{tr}(A^{k-1})\leqslant(\mathrm{tr}(A))^{k-1},$$

在 (6.2.1) 中, 令 $B=A^{k-1}$, 结合上式, 我们有

$$\mathrm{tr}(A^k)\leqslant\mathrm{tr}(A)(\mathrm{tr}(A))^{k-1}\leqslant(\mathrm{tr}(A))^k.$$

推论证毕.

定理 6.2.3 设 A 和 B 为两个 n 阶 Hermite 阵.

(1) 若 $A\geqslant 0$, $B\geqslant 0$. 则

$$\mathrm{tr}(I_n+AB)A\leqslant(1+\lambda_1(AB))\mathrm{tr}A, \tag{6.2.4}$$

$$\mathrm{tr}(I_n+AB)A\leqslant(n+\mathrm{tr}AB)\lambda_1(A); \tag{6.2.5}$$

(2) 若 $A>0$, $B\geqslant 0$, 则

$$\mathrm{tr}(I_n+AB)^{-1}A\geqslant\frac{\mathrm{tr}A}{1+\lambda_1(AB)}. \tag{6.2.6}$$

证明 (1) 因为

$$\mathrm{tr}(I_n+AB)A=\mathrm{tr}A+\mathrm{tr}(ABA)=\mathrm{tr}A+\mathrm{tr}(A^{1/2}BA^{1/2}A), \tag{6.2.7}$$

应用定理 6.2.2, 得

$$\mathrm{tr}(A^{1/2}BA^{1/2}A)\leqslant\lambda_1(A^{1/2}BA^{1/2})\mathrm{tr}A=\lambda_1(AB)\mathrm{tr}A,$$

代入 (6.2.7) 便得到 (6.2.4).

另一方面, 应用定理 6.2.2, 我们有

$$\begin{aligned}\mathrm{tr}(I+AB)A&=\mathrm{tr}(I+A^{1/2}BA^{1/2})A\leqslant\mathrm{tr}(I+A^{1/2}BA^{1/2})\lambda_1(A)\\&=(n+\mathrm{tr}AB)\lambda_1(A).\end{aligned}$$

于是 (6.2.5) 得证.

(2) 因为 $A>0$, 利用 §6.1 迹的性质 (3), 能够验证

$$\operatorname{tr}(I_n+AB)^{-1}A=\operatorname{tr}(I_n+A^{1/2}BA^{1/2})^{-1}A.$$

再利用 (6.2.2) 得

$$\begin{aligned}\operatorname{tr}(I_n+AB)^{-1}A &\geqslant \frac{\operatorname{tr}A}{\lambda_1(I_n+A^{1/2}BA^{1/2})}=\frac{\operatorname{tr}A}{1+\lambda_1(A^{1/2}BA^{1/2})}\\ &=\frac{\operatorname{tr}A}{1+\lambda_1(AB)}.\end{aligned}$$

定理证毕.

定理 6.2.4　设 A 为 n 阶 Hermite 阵, $A\geqslant 0$ 且分块为

$$A=\begin{pmatrix} A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

其中 A_{11} 为 $p\times p$ 方阵, 则

$$\operatorname{tr}A_{21}A_{12}\leqslant \operatorname{tr}A_{11}\cdot\operatorname{tr}A_{22}. \tag{6.2.8}$$

证明　定义

$$A(\varepsilon)=\begin{pmatrix} A_{11}+\varepsilon I_p & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

这里 $\varepsilon>0$. 于是 $A(\varepsilon)>0$. 应用引理 3.2.1(1) 知

$$A_{22}-A_{21}(A_{11}+\varepsilon I_p)^{-1}A_{12}>0.$$

再应用上节性质 (9) 及推论 6.2.2 得到

$$\begin{aligned}\operatorname{tr}A_{22}&>\operatorname{tr}A_{21}(A_{11}+\varepsilon I_p)^{-1}A_{12}=\operatorname{tr}A_{12}A_{21}(A_{11}+\varepsilon I_p)^{-1}\\ &\geqslant\frac{\operatorname{tr}A_{21}A_{12}}{\operatorname{tr}(A_{11}+\varepsilon I_p)},\end{aligned}$$

即

$$\operatorname{tr}A_{22}\cdot\operatorname{tr}(A_{11}+\varepsilon I_p)\geqslant\operatorname{tr}A_{21}A_{12},$$

命 $\varepsilon\to 0$, 即得欲证.

注　当 $n=2, p=1$ 时, 设

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix},$$

则由 $A \geqslant 0$ 可推得 $\det A \geqslant 0$, 于是 $a_{12}a_{21} \leqslant a_{11}a_{22}$. 可见 (6.2.8) 是这个事实的一个推广.

定理 6.2.5 设 A 和 B 为同阶 Hermite 阵, 则

$$2\mathrm{tr}AB \leqslant \mathrm{tr}A^2 + \mathrm{tr}B^2,$$

等号成立 $\Longleftrightarrow A = B$.

证明容易从展开 $(A-B)^2 \geqslant 0$ 并取 tr 来完成.

定理 6.2.6 设 A 和 B 为 n 阶半正定 Hermite 阵, 则对任意两个 n 阶收缩矩阵 U 和 V, 即 $\sigma_1(U) \leqslant 1$ 和 $\sigma_1(V) \leqslant 1$, 有

$$\mathrm{tr}(A-B) \leqslant \mathrm{tr}((A-UBV)^*(A-UBV))^{1/2} \leqslant \mathrm{tr}(A+B), \tag{6.2.9}$$

这里 $\lambda_1(C) \geqslant \cdots \geqslant \lambda_n(C)$ 和 $\sigma_1(C) \geqslant \cdots \geqslant \sigma_n(C)$ 分别为 n 阶矩阵 C 的特征值和奇异值.

证明 我们首先证明若 $A \geqslant 0$, 则对任意两个 n 阶收缩矩阵 U 和 V, 都有

$$\mathrm{Re}(\mathrm{tr}(A-UAV)) \geqslant 0. \tag{6.2.10}$$

由于 $A \geqslant 0$, 故存在一个酉矩阵 P, 使得 $A = P\Lambda P^*$, 其中 $\Lambda = \mathrm{diag}(\lambda_1(A), \cdots, \lambda_n(A))$. 记 $P^*VUP = R = (r_{ij})$, 则 $|r_{ij}| \leqslant 1$, 且

$$\begin{aligned}
\mathrm{Re}(\mathrm{tr}(A-UAV)) &= \mathrm{Re}(\mathrm{tr}(A-AVU)) \\
&= \mathrm{Re}(\mathrm{tr}(\Lambda - \Lambda P^*VUP)) \\
&= \sum_{i=1}^{n} \lambda_i(A)(1-\mathrm{Re}\ r_{ii}) \\
&\geqslant 0.
\end{aligned}$$

容易验证, 当 (6.2.10) 中的减号改为加号时, 不等式仍然成立.

对 $A-UBV$ 进行奇异值分解, 得

$$A-UBV = TDG^*,$$

其中 T 和 G^* 为酉矩阵, $D = \mathrm{diag}\Big(\sigma_1(A-UBV), \cdots, \sigma_n(A-UBV)\Big)$.
于是我们有

$$\Big((A-UAV)^*(A-UAV)\Big)^{1/2} = GDG^* = (GT^*)(A-UAV). \tag{6.2.11}$$

记 $Q = GT^*$, 易证 Q 是酉矩阵, 结合 (6.2.10) 和 (6.2.11), 我们有

$$\begin{aligned}
\mathrm{tr}(A-B) &= \mathrm{Re}(\mathrm{tr}(A-B)) \\
&= \mathrm{Re}(\mathrm{tr}(A-UBV)) - \mathrm{Re}(\mathrm{tr}(B-UBV)) \\
&\leqslant \mathrm{Re}(\mathrm{tr}(A-UBV)) \\
&\leqslant \mathrm{tr}\Big((A-UAV)^*(A-UAV)\Big)^{1/2} \\
&= \mathrm{tr}(QA-QUAV) \\
&= \mathrm{tr}(A+B) - \mathrm{Re}(\mathrm{tr}(A-QA)) - \mathrm{Re}(\mathrm{tr}(B+QUBV)) \\
&\leqslant \mathrm{tr}(A+B).
\end{aligned}$$

定理证毕.

该定理是由 Wang Boying, Xi Boyan 和 Zhang Fuzhen (1999) 提出的. 显然, (6.2.9) 的等价形式为

$$\sum_{i=1}^{n}\lambda_i(A-B) \leqslant \sum_{i=1}^{n}\sigma_i(A-UBV) \leqslant \sum_{i=1}^{n}\lambda_i(A+B). \tag{6.2.12}$$

§6.3　矩阵幂的迹

本节将给出有关 $\mathrm{tr}A^2$, $\mathrm{tr}(A+B)^2$ 和 $\mathrm{tr}(AB)^k$ 的一些不等式.

设 A 的特征值全为实数, 且有 t 个不等于零. 显然 $t>0 \Longleftrightarrow \mathrm{tr}A^2>0$.

定理 6.3.1　设 n 阶方阵 A 的所有特征值都是实数, 且 $\mathrm{tr}A^2>0$.

(1) 若 A 恰有 k 个非零特征值, 则

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2} \leqslant k; \tag{6.3.1}$$

(2) 若 A 恰有 k_1 个正特征值, k_2 个负特征值, 则当 $\mathrm{tr}A \geqslant 0$ 时, 有

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2} \leqslant k_1, \tag{6.3.2}$$

而当 $\mathrm{tr}A \leqslant 0$ 时, 有

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2} \leqslant k_2. \tag{6.3.3}$$

证明　设 A 的 n 个特征值为 $\lambda_1,\cdots,\lambda_n$.

(1) 因为 $\mathrm{tr}A^2>0$, 故 $k>0$. 不妨设 $\lambda_1,\cdots,\lambda_k$ 为非零, 而 $\lambda_{k+1}=\cdots=\lambda_n=0$. 于是 A^2 的特征值 $\lambda_1^2,\cdots,\lambda_k^2$ 不为零, 而其余的特征值 $\lambda_{k+1}^2=\cdots=\lambda_n^2=0$. 考虑如

下平方和

$$H=\sum_{i=1}^{k}(\lambda_i-\overline{\lambda})^2,$$

其中 $\overline{\lambda}=\frac{1}{k}\mathrm{tr}A$. 显然 $H\geqslant 0$ 且 $H=0\Longleftrightarrow\lambda_1=\cdots=\lambda_k=\overline{\lambda}$.

因为

$$H=\sum_{i=1}^{k}\lambda_i^2-k\overline{\lambda}^2=\mathrm{tr}A^2-k\left(\frac{\mathrm{tr}A}{k}\right)^2\geqslant 0.$$

于是 (6.3.1) 得证.

(2) 不妨假设 $\lambda_1,\cdots,\lambda_{k_1}$ 和 $\lambda_{k_1+1},\cdots,\lambda_k$ 分别为 A 的 k_1 个正特征值和 k_2 个负特征值, 这里 $k_1+k_2=k$. 则

$$(\mathrm{tr}A)^2=\left(\sum_{i=1}^{k_1}\lambda_i-\sum_{i=k_1+1}^{k}|\lambda_i|\right)^2,$$

$$\mathrm{tr}A^2=\sum_{i=1}^{k}\lambda_i^2\geqslant\max\left\{\sum_{i=1}^{k_1}\lambda_i^2,\quad\sum_{i=k_1+1}^{k}\lambda_i^2\right\}.$$

当 $\mathrm{tr}A\geqslant 0$ 时, 即 $\sum_{i=1}^{k_1}\lambda_i\geqslant\sum_{i=k_1+1}^{k}|\lambda_i|$, 因此

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2}\leqslant\frac{\left(\sum\limits_{i=1}^{k_1}\lambda_i-\sum\limits_{i=k_1+1}^{k}|\lambda_i|\right)^2}{\sum\limits_{i=1}^{k_1}\lambda_i^2}\leqslant\frac{\left(\sum\limits_{i=1}^{k_1}\lambda_i\right)^2}{\sum\limits_{i=1}^{k_1}\lambda_i^2}\leqslant k_1,$$

上式的最后一个不等式的证明类似于 (1). 事实上, 记 $\overline{\lambda}(k_1)=\sum\limits_{i=1}^{k_1}\lambda_i/k_1$, 于是

$$\sum_{i=1}^{k_1}\left(\lambda_i-\overline{\lambda}(k_1)\right)^2=\sum_{i=1}^{k_1}\lambda_i^2-k_1\overline{\lambda}^2(k_1)\geqslant 0.$$

同理, 当 $\mathrm{tr}A\leqslant 0$ 时, $\sum\limits_{i=1}^{k_1}\lambda_i\leqslant\sum\limits_{i=k_1+1}^{k}|\lambda_i|$, 于是

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2}\leqslant\frac{\left(\sum\limits_{i=k_1+1}^{k}|\lambda_i|-\sum\limits_{i=1}^{k_1}\lambda_i\right)^2}{\sum\limits_{i=1}^{k_1}\lambda_i^2}\leqslant\frac{\left(\sum\limits_{i=k_1+1}^{k}\lambda_i\right)^2}{\sum\limits_{i=k_1+1}^{k}\lambda_i^2}\leqslant k_2.$$

定理证毕.

(3) 在 (6.3.2) 证明中, 用 $-A$ 代替 A, 即得 (6.3.3). 定理证毕.

推论 6.3.1 设 A 的所有特征值都是实数, 正负特征值的个数分别为 k_1 和 k_2, 且 $\mathrm{tr}A^2>0$, 则

$$\frac{(\mathrm{tr}A)^2}{\mathrm{tr}A^2}\leqslant\max(k_1,k_2)\leqslant r(A),\tag{6.3.4}$$

其中 $r(A)$ 表示 A 的秩.

定理 6.3.2 设 A 和 B 为两个同阶 Hermite 阵, 则

$$\mathrm{tr}(AB)^2\leqslant\mathrm{tr}(A^2B^2),$$

且等号成立 $\Longleftrightarrow AB=BA$.

证明 定义 $D=AB-BA$. 则 $D^*=-D$, 即 D 为斜 Hermite 阵. 利用 $\mathrm{tr}AB=\mathrm{tr}BA$, 我们有

$$0\leqslant\mathrm{tr}(DD^*)=2\mathrm{tr}(A^2B^2)-2\mathrm{tr}(AB)^2.$$

定理得证.

注 Marcus(1956) 把本定理结论推广到如下形式.

$$\mathrm{tr}(AB)^k\leqslant\mathrm{tr}(A^kB^k),$$

其中, A 和 B 为同阶 Hermite 阵, k 为正整数.

定理 6.3.3 设 A 和 B 为两个同阶 Hermite 阵, $\mathrm{tr}A>0$, $\mathrm{tr}B>0$. 则

$$\frac{\mathrm{tr}(A+B)^2}{\mathrm{tr}(A+B)}\leqslant\frac{\mathrm{tr}A^2}{\mathrm{tr}A}+\frac{\mathrm{tr}B^2}{\mathrm{tr}B}.\tag{6.3.5}$$

证明 因为 $\mathrm{tr}A>0$, $\mathrm{tr}B>0$, 故 (6.3.5) 等价于

$$2\mathrm{tr}AB\cdot\mathrm{tr}A\cdot\mathrm{tr}B\leqslant(\mathrm{tr}A)^2\mathrm{tr}B^2+(\mathrm{tr}B)^2\mathrm{tr}A^2.\tag{6.3.6}$$

以下证明 (6.3.6). 应用定理 6.2.1 及算术平均与几何平均不等式, 有

$$\begin{aligned}&2\mathrm{tr}AB\cdot\mathrm{tr}A\cdot\mathrm{tr}B\\\leqslant&2(\mathrm{tr}A^2)^{1/2}(\mathrm{tr}B^2)^{1/2}\mathrm{tr}A\cdot\mathrm{tr}B\\=&2\mathrm{tr}A(\mathrm{tr}B^2)^{1/2}\cdot\mathrm{tr}B(\mathrm{tr}A^2)^{1/2}\\\leqslant&[\mathrm{tr}A(\mathrm{tr}B^2)^{1/2}]^2+[\mathrm{tr}B(\mathrm{tr}A^2)^{1/2}]^2\\=&(\mathrm{tr}A)^2\mathrm{tr}B^2+(\mathrm{tr}B)^2\mathrm{tr}A^2,\end{aligned}$$

(6.3.6) 得证. 定理证毕.

Yang(2000) 给出了 $\operatorname{tr}(AB)^k$ 的一个容易计算的上界.

定理 6.3.4 设 A 和 B 为同阶半正定 Hermite 阵, 则对任意的正整数 k, 有

$$\operatorname{tr}(AB)^k \leqslant (\operatorname{tr}A)^k(\operatorname{tr}B)^k.$$

证明 结合 (6.2.1) 和注 6.3.1, 得

$$\operatorname{tr}(AB)^k \leqslant \operatorname{tr}(A^kB^k) \leqslant \operatorname{tr}(A^k)\operatorname{tr}(B^k),$$

由推论 6.2.3, 定理得证.

定理的证明是由本书作者给出的, 比 Yang(2000) 给出的证明更简单.

§6.4 Neumann 不等式及其推广

本节将讨论著名的 Neumann 不等式. 鉴于这个不等式的重要性, 文献中已有许多种证法. 这里给出的证法也许是最简单的. 我们先证明两个引理.

引理 6.4.1(Abel 恒等式) 对任意的两组数 $a_1,\cdots,a_n$ 和 $b_1,\cdots,b_n$, 恒有

$$\sum_{i=1}^{n} a_ib_i = \sum_{i=1}^{n-1}\left[(a_i - a_{i+1})\sum_{j=1}^{i} b_j\right] + a_n\sum_{j=1}^{n} b_j.$$

证明 上式右边可改写为

$$\begin{aligned}
&\sum_{i=1}^{n-1}\left(a_i\sum_{j=1}^{i} b_j\right) - \sum_{i=1}^{n-1}\left(a_{i+1}\sum_{j=1}^{i} b_j\right) + a_n\sum_{j=1}^{n} b_j\\
=&\sum_{i=1}^{n-1}\left(a_i\sum_{j=1}^{i} b_j\right) - \sum_{i=2}^{n}\left(a_i\sum_{j=1}^{i-1} b_j\right) + a_n\sum_{j=1}^{n} b_j\\
=&\sum_{i=2}^{n-1} a_i\left(\sum_{j=1}^{i} b_j - \sum_{j=1}^{i-1} b_j\right) + a_1b_1 - a_n\sum_{j=1}^{n-1} b_j + a_n\sum_{j=1}^{n} b_j\\
=&\sum_{i=2}^{n-1} a_ib_i + a_1b_1 + a_n\left(\sum_{j=1}^{n} b_j - \sum_{j=1}^{n-1} b_j\right)\\
=&\sum_{i=1}^{n} a_ib_i.
\end{aligned}$$

证毕.

下面的引理在定理 4.2.2 中曾经证明过, 这里我们给出另一种证法.

引理 6.4.2　设 A 为 $n\times n$ Hermite 阵, 其特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n$, 对应的标准正交化特征向量为 $\varphi_1,\cdots,\varphi_n$. 设 $x_1,\cdots,x_k$ 为相互正交的向量, 则

$$\sum_{i=1}^{k}\frac{x_i^*Ax_i}{x_i^*x_i}\leqslant\sum_{i=1}^{k}\lambda_i,\qquad k=1,\cdots,n-1,\tag{6.4.1}$$

$$\sum_{i=1}^{n}\frac{x_i^*Ax_i}{x_i^*x_i}\leqslant \mathrm{tr}A,\tag{6.4.2}$$

且当 $x_i\propto\varphi_i$, $i=1,\cdots,n$ 时, 等号成立.

证明　因为 A 为 Hermite 阵, 故存在酉阵 Q, 使得

$$A=Q\Lambda Q^*,\qquad \Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$$

记 P 为 $n\times k$ 阵, 其第 i 列为

$$\frac{Q^*x_i}{(x_i^*x_i)^{1/2}},\qquad i=1,\cdots,k.$$

于是 $P^*P=I_k$. 因而

$$\sum_{i=1}^{k}\frac{x_i^*Ax_i}{x_i^*x_i}=\mathrm{tr}P^*\Lambda P=\mathrm{tr}\Lambda PP^*=\sum_{i=1}^{n}\lambda_i b_{ii}.\tag{6.4.3}$$

这里 b_{ii} 为幂等方阵 $B=(b_{ij})=PP^*$ 的第 i 个对角元, $r(B)=k$. 当 $k=n$ 时, $PP^*=I_n$, 此时所有 $b_{ii}=1$, 于是 (6.4.3) 即为 (6.4.2). 若 $k\leqslant n-1$ 时, 注意到

$$\sum_{i=1}^{n}b_{ii}=\mathrm{tr}B=\mathrm{tr}PP^*=\mathrm{tr}P^*P=k,$$

因为幂等阵的特征值只能为 0 或 1, 应用推论 4.1.2 知: $0\leqslant b_{ii}\leqslant 1$, $i=1,\cdots,n$. 于是

$$\begin{aligned}\sum_{i=1}^{n}\lambda_i b_{ii}&\leqslant\sum_{i=1}^{k}\lambda_i b_{ii}+\left(\sum_{i=k+1}^{n}b_{ii}\right)\lambda_{k+1}\\&=\sum_{i=1}^{k}\lambda_i b_{ii}+\left(\mathrm{tr}B-\sum_{i=1}^{k}b_{ii}\right)\lambda_{k+1}\\&=\sum_{i=1}^{k}\lambda_i b_{ii}+\left(k-\sum_{i=1}^{k}b_{ii}\right)\lambda_{k+1}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} (\lambda_i - \lambda_{k+1}) b_{ii} + k\lambda_{k+1} \\
&\leqslant \sum_{i=1}^{k} (\lambda_i - \lambda_{k+1}) + k\lambda_{k+1} \\
&= \sum_{i=1}^{k} \lambda_i,
\end{aligned}
$$

结合 (6.4.3), 定理得证.

推论 6.4.1 设 $A=(a_{ij})$ 为 n 阶 Hermite 阵, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为其特征值. 则对任意的 $1 \leqslant j_1 \leqslant \cdots \leqslant j_k \leqslant n$ 有

$$
\sum_{i=1}^{k} a_{j_i j_i} \leqslant \sum_{i=1}^{k} \lambda_i, \qquad k = 1, \cdots, n-1. \tag{6.4.4}
$$

证明 在 (6.4.1) 中, 取 $x_i = (0, \cdots, 0, 1, 0, \cdots, 0)'$, 其中 1 位于第 j_i 元位置, $i = 1, \cdots, k$, 即得 (6.4.4).

注 注意在 (6.4.4) 中, 当 $k = n$ 时, 等式成立. 推论 6.4.1 及 $k = n$ 时的等式关系所描述的 Hermite 阵的对角元及特征值的这种关系, 称为受控 (Majorization), 我们将在第八章详细讨论这个问题.

现在我们证明 Neumann 不等式.

定理 6.4.1(Neumann) 设 A 与 B 为两个 n 阶 Hermite 阵, 它们的特征值分别为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 和 $\mu_1 \geqslant \cdots \geqslant \mu_n$. 则

$$
\sum_{i=1}^{n} \lambda_i \mu_{n-i+1} \leqslant \mathrm{tr} AB \leqslant \sum_{i=1}^{n} \lambda_i \mu_i. \tag{6.4.5}
$$

并且左边等式成立 $\Longleftrightarrow B = \sum_{i=1}^{n} \mu_{n-i+1} \varphi_i \varphi_i^*$, 而右边等式成立 $\Longleftrightarrow B = \sum_{i=1}^{n} \mu_i \varphi_i \varphi_i^*$, 这里 $\varphi_1, \cdots, \varphi_n$ 为 A 的对应于 $\lambda_1, \cdots, \lambda_n$ 的标准正交化特征向量.

证明 记 $\boldsymbol{\Phi} = (\varphi_1, \cdots, \varphi_n)$, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 因 A 为 Hermite 阵, 故 A 可表为

$$
A = \boldsymbol{\Phi} \boldsymbol{\Lambda} \boldsymbol{\Phi}^*.
$$

于是

$$
\mathrm{tr} AB = \mathrm{tr}\, \boldsymbol{\Phi} \boldsymbol{\Lambda} \boldsymbol{\Phi}^* B = \mathrm{tr}\, \boldsymbol{\Lambda} \boldsymbol{\Phi}^* B \boldsymbol{\Phi} = \sum_{i=1}^{n} \lambda_i c_{ii},
$$

这里

$$
C = (c_{ij}) = \boldsymbol{\Phi}^* B \boldsymbol{\Phi},
$$

且 $\lambda_i(C) = \lambda_i(B) = \mu_i$, $i = 1, \cdots, n$. 应用引理 6.4.1 和推论 6.4.1, 我们有

$$\begin{aligned}\operatorname{tr}AB &= \sum_{i=1}^{n-1}\left[(\lambda_i - \lambda_{i+1})\sum_{j=1}^{i} c_{jj}\right] + \lambda_n \sum_{j=1}^{n} c_{jj} \\ &\leqslant \sum_{i=1}^{n-1}\left[(\lambda_i - \lambda_{i+1})\sum_{j=1}^{i} \mu_j\right] + \lambda_n \sum_{j=1}^{n} \mu_j \\ &= \sum_{i=1}^{n} \lambda_i \mu_i,\end{aligned}$$

于是右边不等式得证, 其等号成立 $\Longleftrightarrow c_{jj}$ 为 C 的特征值 μ_j, $j = 1, \cdots, n, \Longleftrightarrow$ $\boldsymbol{\Phi}^* B \boldsymbol{\Phi} = \operatorname{diag}(\mu_1, \cdots, \mu_n) \Longleftrightarrow B = \sum\limits_{i=1}^{n} \mu_i \varphi_i \varphi_i^*$.

在右边不等式中, 用 $-B$ 代替 B, 便得到左边不等式及其余结论. 证毕.

当 A 和 B 不是方阵时, 我们也有类似于 (6.4.5) 不等式的结果, 但其中的特征值要用奇异值代替. 这就是下面的推论.

推论 6.4.2　设 A 和 B 皆为 $m \times n$ 矩阵, 它们的奇异值分别为 $\delta_1 \geqslant \cdots \geqslant \delta_n$ 和 $\gamma_1 \geqslant \cdots \geqslant \gamma_n$. 则

$$-\sum_{i=1}^{n} \delta_i \gamma_i \leqslant \operatorname{Re}[\operatorname{tr}AB^*] \leqslant \sum_{i=1}^{n} \delta_i \gamma_i, \tag{6.4.6}$$

其中, 左边等号成立 $\Longleftrightarrow B = -\sum\limits_{i=1}^{n} \gamma_i \varphi_i \psi_i^*$; 而右边等号成立 $\Longleftrightarrow B = \sum\limits_{i=1}^{n} \gamma_i \varphi_i \psi_i^*$, 这里 $\varphi_1, \cdots, \varphi_m$ 和 $\psi_1, \cdots, \psi_n$ 分别 AA^* 和 A^*A 对应于 $\delta_1^2, \cdots, \delta_n^2$ 的标准正交化特征向量.

证明　设 $r(A) = l$, $r(B) = t$, 则

$$\begin{aligned} &\delta_1 \geqslant \cdots \geqslant \delta_l > 0, \qquad \delta_{l+1} = \cdots = \delta_l = 0, \\ &\gamma_1 \geqslant \cdots \geqslant \gamma_t > 0, \qquad \gamma_{t+1} = \cdots = \gamma_n = 0. \end{aligned}$$

定义

$$M_1 = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}.$$

设 λ 为 M_1 的非零特征值, 利用引理 3.2.1(3), 得

$$\det(\lambda I - M_1) = \det\begin{pmatrix} \lambda I_m & -A \\ -A^* & \lambda I_n \end{pmatrix}$$

$$
\begin{aligned}
&= \det(\lambda I_m)\det(\lambda I_n - \lambda^{-1}A^*A)\\
&= \lambda^{m-n}\det(\lambda^2 I - A^*A) = 0.
\end{aligned}
$$

这表明

$$\lambda = \pm\delta_i, \quad i = 1, \cdots, l,$$

于是, M_1 的全部特征值若按递减排列, 则为

$$\delta_1, \cdots, \delta_l, 0, \cdots, 0, -\delta_l, \cdots, -\delta_1, \tag{6.4.7}$$

这里 0 为 $m+n-2l$ 重特征值.

类似地, 定义

$$M_2 = \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}.$$

同理可证, M_2 的全部特征值若按递减排列, 则为

$$\gamma_1, \cdots, \gamma_t, 0, \cdots, 0, -\gamma_t, \cdots, -\gamma_1.$$

因为

$$
\begin{aligned}
\mathrm{tr}M_1M_2 &= \mathrm{tr}\begin{pmatrix} AB^* & 0 \\ 0 & A^*A \end{pmatrix} = \mathrm{tr}AB^* + \mathrm{tr}A^*B\\
&= \mathrm{tr}AB^* + \overline{\mathrm{tr}AB^*} = 2\mathrm{Re}[\mathrm{tr}AB^*].
\end{aligned}
$$

应用定理 6.4.1, 得

$$
\begin{aligned}
2\mathrm{Re}[\mathrm{tr}AB^*] = \mathrm{tr}M_1M_2 &\leqslant \sum_{i=1}^{\min(l,t)} \delta_i\gamma_i + \sum_{i=1}^{\min(l,t)} (-\delta_i)(-\gamma_i)\\
&= 2\sum_{i=1}^{n} \delta_i\gamma_i, \qquad (6.4.8)\\
2\mathrm{Re}[\mathrm{tr}AB^*] = \mathrm{tr}M_1M_2 &\geqslant \sum_{i=1}^{\min(l,t)} \delta_i(-\gamma_i) + \sum_{i=1}^{\min(l,t)} (-\delta_i)\gamma_i\\
&= -2\sum_{i=1}^{n} \delta_i\gamma_i,
\end{aligned}
$$

这就证明了 (6.4.6).

容易验证, M_1 对应于 (6.4.7) 的 $m+n$ 个特征值的 $m+n$ 个线性无关的特征向量依次为

$$g_1=\begin{pmatrix}\varphi_1\\ \psi_1\end{pmatrix},\cdots,\quad g_l=\begin{pmatrix}\varphi_l\\ \psi_l\end{pmatrix},\quad g_{l+1}=\begin{pmatrix}\varphi_{l+1}\\ 0\end{pmatrix},\cdots,g_{m-l}=\begin{pmatrix}\varphi_m\\ 0\end{pmatrix},$$

$$g_{m-l+1}=\begin{pmatrix}0\\ \psi_{l+1}\end{pmatrix},\cdots,\ g_{m+n-l}=\begin{pmatrix}0\\ \psi_n\end{pmatrix},g_{m+n-l+1}=\begin{pmatrix}-\varphi_1\\ \psi_1\end{pmatrix},\cdots,\ g_{m+n}=\begin{pmatrix}-\varphi_l\\ \psi_l\end{pmatrix},$$

又根据定理 6.4.1 知,(6.4.8) 中等号成立, 当且仅当

$$\begin{aligned}M_2=\sum_{i=1}^{m+n}\lambda_i(M_2)g_ig_i^*&=\sum_{i=1}^{t}\gamma_i\begin{pmatrix}\varphi_i\\ \psi_i\end{pmatrix}(\varphi_i^*,\psi_i^*)\\&+\sum_{i=1}^{t}(-\gamma_i)\begin{pmatrix}-\varphi_i\\ \psi_i\end{pmatrix}(-\varphi_i^*,\psi_i^*)\\&=\begin{pmatrix}0 & 2\sum\limits_{i=1}^{t}\gamma_i\varphi_i\psi_i^*\\ 2\sum\limits_{i=1}^{t}\gamma_i\psi_i\varphi_i^* & 0\end{pmatrix}.\end{aligned}$$

这等价于 (6.4.6) 的右边等号成立 $\Longleftrightarrow B=\sum\limits_{i=1}^{n}\gamma_i\varphi_i\psi_i^*$. 基于此, 其余结论的证明是显然的. 定理证毕.

推论 6.4.3　采用推论 6.4.2 的记号, 若 A 和 B 为实矩阵, 则

$$-\sum_{i=1}^{n}\delta_i\gamma_i\leqslant \mathrm{tr}AB'\leqslant\sum_{i=1}^{n}\delta_i\gamma_i,\tag{6.4.9}$$

且左边等号成立 $\Longleftrightarrow B=-\sum\limits_{i=1}^{n}\gamma_i\varphi_i\psi_i'$, 而右边等号成立 $\Longleftrightarrow B=\sum\limits_{i=1}^{n}\gamma_i\varphi_i\psi_i'$. 其中 $\varphi_1,\cdots,\varphi_n$ 和 $\psi_1,\cdots,\psi_n$ 分别为 AA' 和 $A'A$ 对应于 $\delta_1^2,\cdots,\delta_n^2$ 的标准正交化特征向量.

不等式 (6.4.6) 和 (6.4.9) 也常常称作 Neumann 不等式.

为了对 Neumann 不等式作一些推广, 我们需要如下引理.

引理 6.4.3 假设 $\alpha_1 \geqslant \cdots \geqslant \alpha_n$, $\mu_1 \geqslant \cdots \geqslant \mu_n$. 若 $\pi(1),\cdots,\pi(n)$ 为 $1,\cdots,n$ 的任一排列, 则

$$\sum_{i=1}^{n} \alpha_i \mu_{n-i+1} \leqslant \sum_{i=1}^{n} \alpha_{\pi(i)} \mu_i \leqslant \sum_{i=1}^{n} \alpha_i \mu_i.$$

证明 我们只需证明右边不等式. 因为对 $\alpha_1 \geqslant \cdots \geqslant \alpha_n$ 和 $-\mu_n \geqslant \cdots \geqslant -\mu_1$ 应用右边不等式, 即得左边的不等式.

假设对 $i=1,\cdots,k$, 有 $\pi(i)=i$, 但是 $\pi(k+1)=k+t$, $t>1$, 而 $\pi(k+s)=k+1$, $s>1$. 现在我们把和式 $\sum_i \alpha_{\pi(i)}\mu_i$ 中的两项 $\alpha_{\pi(k+1)}\mu_{k+1}$ 和 $\alpha_{\pi(k+s)}\mu_{k+s}$ 分别换成 $\alpha_{\pi(k+s)}\mu_{k+1}$ 和 $\alpha_{\pi(k+1)}\mu_{k+s}$. 注意到

$$\begin{aligned}
&(\alpha_{\pi(k+s)}\mu_{k+1} + \alpha_{\pi(k+1)}\mu_{k+s}) - (\alpha_{\pi(k+1)}\mu_{k+1} + \alpha_{\pi(k+s)}\mu_{k+s}) \\
=&\alpha_{k+1}\mu_{k+1} + \alpha_{k+t}\mu_{k+s} - \alpha_{k+t}\mu_{k+1} - \alpha_{k+1}\mu_{k+s} \\
=&(\alpha_{k+1} - \alpha_{k+t})(\mu_{k+1} - \mu_{k+s}) \geqslant 0,
\end{aligned}$$

于是

$$\sum_{i=k+1}^{n} \alpha_{\pi(i)}\mu_i \leqslant \alpha_{k+1}\mu_{k+1} + \sum_{i=k+2}^{n} \alpha_{\pi'(i)}\mu_i,$$

其中 $\pi'(k+2),\cdots,\pi'(n)$ 是 $k+2,\cdots,n$ 的一个排列. 下一步对和式 $\sum_{i=k+2}^{n} \alpha_{\pi'(i)}\mu_i$ 重复前面的方法, 便可得到所要证明的结论.

现在我们证明 Neumann 不等式的一个推广形式, 它是由 Bushell 和 Trustrum (1990) 证明的.

定理 6.4.2(Bushell-Trustrum) 设 A 和 B 为两个 n 阶半正定 Hermite 阵, 它们的特征值分别为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ 和 $\mu_1 \geqslant \cdots \geqslant \mu_n \geqslant 0$. 则对任意正整数 k, 有

$$\sum_{i=1}^{n} \lambda_i^k \mu_{n-i+1}^k \leqslant \operatorname{tr}(AB)^k \leqslant \sum_{i=1}^{n} \lambda_i^k \mu_i^k. \tag{6.4.10}$$

证明 我们可以假设 $A>0$, $B>0$. 不然的话, 对任意的 $c>0$, 必有 $A+cI>0$, $B+cI>0$, 最后对所得结果命 $c\to 0$ 取极限, 即得欲证.

首先注意到, 存在两个 n 阶酉阵 U_1 和 U_2, 使得对一切 n 阶酉阵 U

$$\operatorname{tr}(AU_2BU_2^*)^k \leqslant \operatorname{tr}(AUBU^*)^k \leqslant \operatorname{tr}(AU_1BU_1^*)^k. \tag{6.4.11}$$

之所以如此, 是因为全体 n 阶酉阵构成一个紧致集合, 而映射 $U\to\operatorname{tr}(AUBU^*)^k$ 为定义在该集合上的连续函数, 所以必然在某个 U_1 和 U_2 上达到最大值和最小值.

在 (6.4.11) 中, 特别取 $U=I$, 则得到

$$\operatorname{tr}(AU_2BU_2^*)^k \leqslant \operatorname{tr}(AB)^k \leqslant \operatorname{tr}(AU_1BU_1^*)^k. \tag{6.4.12}$$

若记 $B_i = U_iBU_i^*(i=1,2)$, 我们将先证明: 存在 n 阶酉阵 X, 使得 X^*AX 和 X^*B_2X 成对角形.

记

$$R=\begin{bmatrix} (1+|\varepsilon|^2)^{-1/2} & -\varepsilon(1+|\varepsilon|^2)^{-1/2} & \underset{1\times(n-2)}{\mathbf{0}} \\ \bar{\varepsilon}(1+|\varepsilon|^2)^{-1/2} & (1+|\varepsilon|^2)^{-1/2} & \underset{1\times(n-2)}{\mathbf{0}} \\ \underset{(n-2)\times 1}{\mathbf{0}} & \underset{(n-2)\times 1}{\mathbf{0}} & I_{n-2} \end{bmatrix},$$

$$F=\begin{bmatrix} 0 & -\dfrac{\varepsilon}{|\varepsilon|} & \underset{1\times(n-2)}{\mathbf{0}} \\ \dfrac{\bar{\varepsilon}}{|\varepsilon|} & 0 & \underset{1\times(n-2)}{\mathbf{0}} \\ \underset{(n-2)\times 1}{0} & \underset{(n-2)\times 1}{0} & \underset{(n-2)\times(n-2)}{\mathbf{0}} \end{bmatrix},$$

其中 I_{n-2} 和 O_{n-2} 分别为 $n-2$ 阶单位阵和零阵. $O_{m\times n}$ 表示 $m\times n$ 阶零阵. 注意, 这里 R 为一酉阵, 且对充分小的 $\varepsilon\neq 0$, R 可表为

$$R = I_n + |\varepsilon|F + O(|\varepsilon|^2).$$

对任意的酉阵 T, 定义

$$\tilde{B} = (TRT^*)B(TR^*T^*) = B + |\varepsilon|T(FC-CF)T^* + O(|\varepsilon|^2), \tag{6.4.13}$$

其中 $C=T^*BT$.

容易验证, 对任意两个 n 阶酉阵 P 和 Q, 有

$$\operatorname{tr}(P+|\varepsilon|Q)^k = \operatorname{tr}P^k + k|\varepsilon|\operatorname{tr}(P^{k-1}Q) + O(|\varepsilon|^2),$$

因而

$$\operatorname{tr}(A\tilde{B})^k - \operatorname{tr}(AB)^k = k|\varepsilon|\operatorname{tr}[D(FC-CF)] + O(|\varepsilon|^2), \tag{6.4.14}$$

这里 $D=T^*(AB)^{k-1}AT$.

因 $(AB)^{k-1}A\geqslant 0$, 故我们可以选择酉阵 T, 使得 D 成对角形. 即

$$\boldsymbol{D} = \operatorname{diag}(d_1,\cdots,d_n), d_1\geqslant\cdots\geqslant d_n\geqslant 0.$$

代入 (6.4.14), 并注意到 $C=(c_{ij})=T^*BT$ 为 Hermite 阵, 则

$$\begin{aligned}\operatorname{tr}(A\tilde{B})^k - \operatorname{tr}(AB)^k &= k(d_{n-1}-d_n)(\bar{\varepsilon}c_{12}+\varepsilon c_{21}) + O(|\varepsilon|^2) \\ &= k(d_{n-1}-d_n)(\bar{\varepsilon}c_{12}+\varepsilon\bar{c}_{21}) + O(|\varepsilon|^2).\end{aligned}$$

若 $d_{n-1} \neq d_n$, 则在上式中, 令 $B = B_i$, $i = 1, 2$, 由 B_i 的定义及 (6.4.12), 我们推得 $\bar{\varepsilon} c_{12} + \varepsilon \bar{c}_{21} = 0$. 若取 $\varepsilon = \eta c_{12}$, $\eta > 0$, 则 $|c_{12}| = 0$, 于是 $c_{12} = \bar{c}_{21} = 0$. 类似地, 若 $d_i \neq d_j$, 用同样的方法可以证明 $c_{ij} = \bar{c}_{ji} = 0$.

假设 $c_1 > c_2 > \cdots > c_l$ 是 $d_1, \cdots, d_n$ 的 l 个不同值, 此时 $D = \mathrm{diag}\,(c_1 I_{n1}, \cdots, c_l I_{nl})$. 把 $C = T^* B T$ 作对应分块

$$C = \mathrm{diag}(C_1, \cdots, C_l),$$

这里 C_i 为 n_i 阶半正定 Hermite 阵. 设 $U_i (i = 1, \cdots, l)$ 为酉阵, 使得 $E_i = U_i^* C_i U_i$, $i = 1, \cdots, l$ 成对角阵. 记

$$U = \mathrm{diag}(U_1, \cdots, U_l),$$

$$E = \mathrm{diag}(E_1, \cdots, E_l).$$

取 $X = TU$, 则 X 为一酉阵, 且当 B 为 B_1 或 B_2 时,

$$X^* B X = U^* T^* B T U = U^* C U = E, \tag{6.4.15}$$

这是一个对角阵, 其对角元为 B 的特征值, 而且

$$X^* (AB)^{k-1} A X = U^* T^* (AB)^{k-1} A T U = U^* D U = D, \tag{6.4.16}$$

最后一个等式是因为 U 和 D 有相同大小的对角块. 由 (6.4.15) 和 (6.4.16) 两式知

$$(E^{1/2} (X^* A X) E^{1/2})^k = E^{1/2} X^* (AB)^{k-1} A X E^{1/2} = E^{1/2} D E^{1/2}.$$

于是

$$X^* A X = E^{-1/2} (E^{1/2} D E^{1/2})^{1/k} E^{-1/2} \tag{6.4.17}$$

是一个对角阵. 这就证明了, $X^* B_1 X$, $X^* B_2 X$ 和 $X^* A X$ 为对角阵.

根据 (6.4.15) 和 (6.4.17), 当 B 为 B_1 或 B_2 时, $X^* B X$ 和 $X^* A X$ 都为对角阵. 但注意到 $X^* B_1 X = X^* U_1 B U_1^* X$, $X^* B_2 X = X^* U_2 B U_2^* X$, U_1, U_2 和 X 皆为酉阵, 所以 $X^* B_i X$ 的对角元是 B 的特征值. 于是

$$\mathrm{tr}(AB)^k = \mathrm{tr}(X^* (AB)^k X) = \mathrm{tr}(X^* A X \; X^* B X)^k$$

$$= \sum_{i=1}^{n} \lambda_{\pi(i)}^k \mu_i^k,$$

由引理 6.4.3, (6.4.10) 得证. 定理证毕.

Neumann 不等式的重要应用之一是矩阵逼近, 这将在下节讨论.

§6.5 矩阵逼近

本节要讨论的矩阵逼近的问题是, 对给定的秩为 t 的 $m \times n$ 矩阵 A, 找一个秩小于 t 的矩阵 A, 使得 $A-B$ 的欧氏范数 $\|A-B\|_{\mathrm{F}} = \mathrm{tr}(A-B)^* \ (A-B)$ 达到最小. 下面我们应用 Neumann 不等式导出该问题的解, 我们只限于讨论实矩阵逼近问题.

定理 6.5.1 设 A 为 $m \times n$ 实矩阵, $r(A)=t$, 其非零奇异值为 $\sigma_1 \geqslant \cdots \geqslant \sigma_t > 0$. 则对任意秩为 $k(k<t)$ 的 $m \times n$ 实矩阵 B, 有

$$\mathrm{tr}(A-B)'(A-B) \geqslant \sum_{i=k+1}^{t} \sigma_i^2, \tag{6.5.1}$$

等号成立 $\Longleftrightarrow B = \sum_{i=1}^{k} \sigma_i \varphi_i \psi_i'$, 这里 $\varphi_1, \cdots, \varphi_k$ 和 $\psi_1, \cdots, \psi_k$ 分别为 AA' 和 $A'A$ 的对应于特征值 $\sigma_1^2, \cdots, \sigma_k^2$ 的标准正交化特征向量.

证明 设 B 的非零奇异值为 $r_1, \cdots, r_k$, 应用 Neumann 不等式 (推论 6.4.3) 得

$$\begin{aligned}
\mathrm{tr}(A-B)'(A-B) &= \mathrm{tr}A'A + \mathrm{tr}B'B - 2\mathrm{tr}A'B \\
&\geqslant \sum_{i=1}^{t} \sigma_i^2 + \sum_{i=1}^{k} \gamma_i^2 - 2\sum_{i=1}^{k} \sigma_i \gamma_i \\
&= \sum_{i=1}^{k} (\sigma_i - \gamma_i)^2 + \sum_{i=k+1}^{t} \sigma_i^2 \\
&\geqslant \sum_{i=k+1}^{t} \sigma_i^2,
\end{aligned}$$

等号成立 $\Longleftrightarrow B = \sum_{i=1}^{k} \sigma_i \varphi_i \psi_i'$. 证毕.

这个结果表明, 若 A 有奇异值分解

$$A = \sum_{i=1}^{t} \sigma_i \varphi_i \psi_i',$$

那么 A 的具有秩 $k(k<t)$ 的最佳逼近为

$$B = \sum_{i=1}^{k} \sigma_i \varphi_i \psi_i',$$

这个事实有许多重要应用. 例如, 在多元统计分析中, 主成分分析和因子分析的数学本质就是矩阵逼近. 本定理正是为这些分析方法提供了理论依据. (参阅王学仁和王松桂 (1990, p. 304)).

特别, 当 A 为实对称半正定阵时, 定理简化为

推论 6.5.1 设 A 为 $n\times n$ 实对称半正定阵, $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 A 的特征值, $r(A)=t$, B 为 $n\times n$ 阵, $r(B)=k$, $(k<t)$. 则

$$\operatorname{tr}(A-B)'(A-B)\geqslant\sum_{i=k+1}^{t}\lambda_i^2.$$

等号成立 $\Longleftrightarrow B=\sum_{i=1}^{k}\lambda_i\varphi_i\varphi_i'$, 这里, $\varphi_1,\cdots,\varphi_k$ 为 A 的对应于 $\lambda_1,\cdots,\lambda_k$ 的标准正交化特征向量.

定理 6.5.1 能够很容易推广到加权矩阵逼近的情形, 即求 B, 使得 $\operatorname{tr}(A-B)'V(A-B)$ 达到最小, 这里 V 为实对称正定阵. 事实上, 用 $V^{1/2}A$ 和 $V^{1/2}B$ 代替定理中的 A 和 B, 我们立即得到如下推论.

推论 6.5.2 设 A 为 $m\times n$ 的实矩阵, $r(A)=t$, V 为 $m\times m$ 实对称正定阵. 记 $A'VA$ 的非零特征值为 $\sigma_1^2\geqslant\cdots\geqslant\sigma_t^2>0$, 则对任意秩为 $k(k<t)$ 的 $m\times n$ 矩阵 B

$$\operatorname{tr}(A-B)'V(A-B)\geqslant\sum_{i=k+1}^{t}\sigma_i^2.$$

等号成立 $\Longleftrightarrow B'VB=\sum_{i=1}^{k}\sigma_i^2\psi_i\psi_i'$, 这里, $\psi_1,\cdots,\psi_k$ 为 $A'VA$ 的对应于特征值 $\sigma_1^2,\cdots,\sigma_k^2$ 的标准正交化特征向量.

最后我们考虑用正交阵去逼近一个实矩阵.

定理 6.5.2 设 A 为 $n\times n$ 实矩阵, 其奇异值为 $\sigma_1\geqslant\cdots\geqslant\sigma_n$. 设 M 为 $n\times n$ 正交阵. 则

$$\operatorname{tr}(A-M)'(A-M)\geqslant\sum_{i=1}^{n}(\sigma_i-1)^2,$$

等号成立 $\Longleftrightarrow M=\sum_{i=1}^{n}\varphi_i\psi_i'$, 这里 $\varphi_1,\cdots,\varphi_n$ 和 $\psi_1,\cdots,\psi_n$ 分别为 AA' 和 $A'A$ 的标准正交化特征向量.

证明 因为正交阵的奇异值全为 1, 由推论 6.4.3 得

$$\operatorname{tr}A'M\leqslant\sum_{i=1}^{p}\sigma_i.$$

于是

$$\begin{aligned}\operatorname{tr}(A-M)'(A-M)&=\operatorname{tr}A'A+n-2\operatorname{tr}A'M\\&\geqslant\sum_{i=1}^{n}\sigma_i^2+n-2\sum_{i=1}^{n}\sigma_i=\sum_{i=1}^{n}(\sigma_i-1)^2,\end{aligned}$$

等号成立 $\Longleftrightarrow M=\sum_{i=1}^{n}\varphi_i\psi_i'$. 证毕.

这个定理表明, 若 A 有奇异值分解

$$A=\sum_{i=1}^{n}\sigma_i\varphi_i\psi_i',$$

则它的最佳正交逼近为

$$M=\sum_{i=1}^{n}\varphi_i\psi_i'.$$

特别, 当 A 为实对称矩阵时, 其最佳正交逼近为单位阵.

§6.6 带约束条件的矩阵迹

假设 A 为 $n\times n$ Hermite 阵, X 为 $n\times p$ 矩阵, 本节我们首先建立在约束条件 $X^*X=I_p$ 下, $\mathrm{tr}X^*AX$ 和 $\mathrm{tr}(X^*AX)^{-1}$ 的一些不等式. 而后对实对称正定阵 A 及实矩阵 X, 在同样的约束条件下给出了 $\mathrm{tr}(X'AX-(X'A^{-1}X)^{-1})$, $\mathrm{tr}(X'AX\cdot X'A^{-1}X)$ 和 $\mathrm{tr}X'AX/\mathrm{tr}(X'A^{-1}X)^{-1}$ 的上界.

定理 6.6.1 设 A 为 $n\times n$ Hermite 阵, $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为其特征值, X 为 $n\times p$ 矩阵且满足 $X^*X=I_p$, 则

$$\sum_{i=n-p+1}^{n}\lambda_i\leqslant\mathrm{tr}X^*AX\leqslant\sum_{i=1}^{p}\lambda_i. \tag{6.6.1}$$

当 $X=(\varphi_{n-p+1},\cdots,\varphi_n)$ 时, 左边等号成立; 当 $X=(\varphi_1,\cdots,\varphi_p)$ 时, 右边等号成立, 其中 $\varphi_1,\cdots,\varphi_n$ 为对应于 $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 的 A 的标准正交化特征向量.

证明 因为 $X^*X=I_p$, 依 Poincaré分离定理 (见定理 4.6.1), 得

$$\lambda_{n-p+i}=\lambda_{n-p+i}(A)\leqslant\lambda_i(X^*AX)\leqslant\lambda_i(A)=\lambda_i,\qquad i=1,\cdots,p. \tag{6.6.2}$$

于是

$$\sum_{i=1}^{p}\lambda_{n-p+i}\leqslant\mathrm{tr}(X^*AX)\leqslant\sum_{i=1}^{p}\lambda_i,$$

此即 (6.6.1), 其余结论容易验证, 定理证毕.

定理 6.6.2 设 A 为 $n\times n$ 正定 Hermite 阵, 其余记号同定理 6.6.1. 则对一切满足 $X^*X=I_p$ 的矩阵 X, 有

$$\sum_{i=1}^{p}\lambda_i^{-1}\leqslant\mathrm{tr}(X^*AX)^{-1}\leqslant\sum_{i=1}^{p}\lambda_{n-p+i}^{-1}. \tag{6.6.3}$$

当 $X=(\varphi_1,\cdots,\varphi_p)$ 时, 左边等号成立, 而当 $X=(\varphi_{n-p+1},\cdots,\varphi_n)$ 时, 右边等号成立.

证明 由 (6.6.2), 我们有

$$\lambda_i^{-1}\leqslant\lambda_i^{-1}(X^*AX)\leqslant\lambda_{n-p+i}^{-1},\qquad i=1,\cdots,p.$$

于是

$$\sum_{i=1}^p\lambda_i^{-1}\leqslant\sum_{i=1}^p\lambda_i^{-1}(X^*AX)\leqslant\sum_{i=1}^p\lambda_{n-p+i}^{-1}.\tag{6.6.4}$$

另一方面, 因为

$$\lambda_i((X^*AX)^{-1})=\frac{1}{\lambda_{p-i+1}(X^*AX)},\qquad i=1,\cdots,p.$$

我们得到

$$\begin{aligned}\operatorname{tr}(X^*AX)^{-1}&=\sum_{i=1}^p\lambda_i((X^*AX)^{-1})=\sum_{i=1}^p\lambda_{p-i+1}^{-1}(X^*AX)\\&=\sum_{i=1}^p\lambda_i^{-1}(X^*AX),\end{aligned}\tag{6.6.5}$$

结合 (6.6.4) 和 (6.6.5), 我们得到 (6.6.3), 其余结论很容易验证. 定理证毕.

在下面的定理中, 所涉及的矩阵都假定是实矩阵.

定理 6.6.3 设 A 为 $n\times n$ 实对称正定阵, $\lambda_1>\cdots>\lambda_n$ 为其特征值, $n>2p$. 则对一切满足 $X'X=I_p$ 的矩阵, 有

$$0\leqslant\operatorname{tr}[X'AX-(X'A^{-1}X)^{-1}]\leqslant\sum_{i=1}^p(\lambda_i^{1/2}-\lambda_{n-i+1}^{1/2})^2.\tag{6.6.6}$$

证明 设 Q 为正交阵, 使得

$$A=Q\Lambda Q',\quad \Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n),$$

若命 $\tilde X=Q'X$, 则 $\tilde X'\tilde X=X'QQ'X=I_p$, 且

$$\operatorname{tr}[X'AX-(X'A^{-1}X)^{-1}]=\operatorname{tr}[\tilde X'\Lambda\tilde X-(\tilde X'\Lambda^{-1}\tilde X)^{-1}],$$

可见, 我们可以假设 (6.6.6) 中的 A 就是对角阵 Λ. 于是以下我们证明: 对一切满足 $X'X=I_p$ 的 X, 有

$$0\leqslant\operatorname{tr}[X'\Lambda X-(X'\Lambda^{-1}X)^{-1}]\leqslant\sum_{i=1}^p(\lambda_i^{1/2}-\lambda_{n-i+1}^{1/2})^2.\tag{6.6.7}$$

我们采用 Lagrange 乘子法. 设 $L=(l_{ij})$ 为由 Lagrange 乘子组成的 $p\times p$ 对称阵. 考虑辅助函数

$$F(X,L)=\mathrm{tr}[X'\Lambda X-(X'\Lambda^{-1}X)^{-1}]-\mathrm{tr}[L(X'X-I_p)],$$

利用 §1.11 的例 1.11.5 和例 1.11.6 得

$$\frac{\partial \mathrm{tr}X'\Lambda X}{\partial X}=2\Lambda X,$$

$$\frac{\partial \mathrm{tr}(X'\Lambda^{-1}X)^{-1}}{\partial X}=-2\Lambda^{-1}X(X'\Lambda^{-1}X)^{-2},$$

$$\frac{\partial \mathrm{tr}LX'X}{\partial X}=2XL.$$

对 $F(X,L)$ 关于 X 求导数, 并令其等于零, 得到

$$X'\Lambda+(X'\Lambda^{-1}X)^{-2}X'\Lambda^{-1}=LX'. \tag{6.6.8}$$

右乘 X 后得到

$$X'\Lambda X+(X'\Lambda^{-1}X)^{-1}=L, \tag{6.6.9}$$

将 (6.6.9) 代入 (6.6.8), 我们有

$$X'\Lambda+(X'\Lambda^{-1}X)^{-2}X'\Lambda^{-1}=[X'\Lambda X+(X'\Lambda^{-1}X)^{-1}]X'.$$

再将上式两边右乘 ΛX, 得

$$X'\Lambda^2X+(X'\Lambda^{-1}X)^{-2}=(X'\Lambda X)^2+(X'\Lambda^{-1}X)^{-1}(X'\Lambda X). \tag{6.6.10}$$

因为 (6.6.10) 的左端和右端第一项都是对称阵, 因而右端第二项的两个矩阵 $(X'\Lambda^{-1}X)^{-1}$ 和 $X'\Lambda X$ 是可交换的, 由此可推知 $X'\Lambda^{-1}X$ 和 $X'\Lambda X$ 也是可交换的. 依据定理 1.3.5, 存在正交阵 $\boldsymbol{\Phi}$, 使得这两个矩阵同时对角化, 即

$$X'\Lambda X=\boldsymbol{\Phi}D\boldsymbol{\Phi}',\quad D=\mathrm{diag}(d_1,\cdots,d_p), \tag{6.6.11}$$

$$X'\Lambda^{-1}X=\boldsymbol{\Phi}\Delta\boldsymbol{\Phi}',\quad \Delta=\mathrm{diag}(\delta_1,\cdots,\delta_p), \tag{6.6.12}$$

将这两个分解式代入 (6.6.10), 再左乘 $\boldsymbol{\Phi}'$, 然后转置所得方程之两边, 最后得到

$$\Lambda W+\Lambda^{-1}W\Delta^{-2}=W(D+\Delta^{-1}), \tag{6.6.13}$$

这里 $W=X\boldsymbol{\Phi}$ 为 $n\times p$ 矩阵. 记 W 的第 j 列为 $(w_1,\cdots,w_n)'$, 则 (6.6.13) 蕴含着

$$\lambda_iw_i+\lambda_i^{-1}w_i\delta_j^{-2}=w_i(d_j+\delta_j^{-1}),\quad i=1,\cdots,n, \tag{6.6.14}$$

等价地

$$w_i[\lambda_i^2 - \lambda_i(d_j + \delta_j^{-1}) + \delta_j^{-2}] = 0, \quad i = 1, \cdots, n. \tag{6.6.15}$$

上式方括号内是关于 λ_i 的一元二次方程, 它最多有两个不同的根, 因而在 (6.6.15) 中, 至多有两个 w_i 不等于零. 首先假设有两个 w_i 不为零, 并设它们是 w_r 和 w_s. 从方程 (6.6.14) 知, 方程

$$\lambda + \lambda^{-1}\delta_j^{-2} = d_j + \delta_j^{-1}$$

有两个根 λ_r 和 λ_s, 根据一元二次方程的根与系数的关系, 我们得到

$$\begin{cases} \lambda_r + \lambda_s = d_j + \delta_j^{-1}, \\ \lambda_r \lambda_s = \delta_j^{-2}. \end{cases}$$

由此可得到

$$d_j - \delta_j^{-1} = (\lambda_r^{1/2} - \lambda_s^{1/2})^2. \tag{6.6.16}$$

如果只有一个 w_i 不为零, 则容易推出

$$d_j - \delta_j^{-1} = 0. \tag{6.6.17}$$

从 (6.6.11) 和 (6.6.12) 我们有

$$\text{tr}[X'\Lambda X - (X'\Lambda^{-1}X)^{-1}] = \text{tr}(D - \Delta^{-1}) = \sum_{j=1}^{p}(d_j - \delta_j^{-1}). \tag{6.6.18}$$

根据上面的讨论, 在 (6.6.18) 右端的和式中, 每一项的值或者为零, 或者为 $(\lambda_r^{1/2} - \lambda_s^{1/2})^2$.

记 $W = X\boldsymbol{\Phi} = (w_{ij})$. 我们上面已经证明, W 的每个列至多有两个非零元素, 且若 W 的第 j 列的两个元素 w_{rj} 和 w_{sj} 不为零, 则 (6.6.18) 右端和式中就有一项为 $(\lambda_r^{1/2} - \lambda_s^{1/2})^2$. 现在我们要证明, 和式中所含的特征值不会重复. 这归结为证明, 在 W 的任一行只有一个非零元素, 即对一切 $i \neq j$,

$$w_{ki}w_{kj} = 0, \qquad k = 1, \cdots, n. \tag{6.6.19}$$

注意到, 由 (6.6.11) 和 $W'W = I_p$ 可推出

$$\sum_{k=1}^{n} \lambda_k w_{ki} w_{kj} = 0, \qquad i \neq j, \tag{6.6.20}$$

$$\sum_{k=1}^{n} w_{ki} w_{kj} = 0, \qquad i \neq j. \tag{6.6.21}$$

因而 W 不可能只有一行有两个非零元素. 事实上, 若只在 t_1 行有两个非零元素 w_{t_1i} 和 $w_{t_1j}, i\neq j$, 则由 (6.6.21) 得 $w_{t_1i}w_{t_1j}=0$, 这是不可能的. 如果 W 有两行具有两个非零元素, 且它们在两个列里, 设为 w_{t_1i}, w_{t_1j} 和 w_{t_2i}, w_{t_2j}. 于是由 (6.6.20) 和 (6.6.21) 得

$$\lambda_{t_1}w_{t_1i}w_{t_1j}+\lambda_{t_2}w_{t_2i}w_{t_2j}=0,$$

$$w_{t_1i}w_{t_1j}+w_{t_2i}w_{t_2j}=0.$$

因为 $\lambda_{t_1}\neq\lambda_{t_2}$, 所以, $w_{t_1i}w_{t_1j}=0$, 矛盾. 类似地讨论可以证明 W 也不可能有两个以上的行有两个非零元素. 这就证明了 (6.6.19).

既然和式中诸 λ_i 不重复出现, 可见最大值为

$$\sum_{i=1}^{p}(\lambda_i^{1/2}-\lambda_{n-i+1}^{1/2})^2.$$

定理证毕.

推论 6.6.1 在定理 6.6.3 中, 若 $p=1$, 则对一切满足 $x'x=1$ 的向量 x, 有

$$0\leqslant x'Ax-(x'A^{-1}x)^{-1}\leqslant(\sqrt{\lambda_1}-\sqrt{\lambda_n})^2.$$

注 1 定理 6.6.3 渊源于线性模型中最小二乘估计相对效率的研究, 关于这一点, 我们将在 §9.2 给予解释.

应用类似的方法, 可以证明如下两个定理. 详细证明分别见杨虎 (1988) 和 Khatri 和 Rao(1981).

定理 6.6.4 在定理 6.6.3 条件下, 有

$$\frac{\mathrm{tr}X'AX}{\mathrm{tr}(X'A^{-1}X)^{-1}}\leqslant\left(\frac{\sum\limits_{i=1}^{p}(\lambda_i+\lambda_{n-i+1})}{2\sum\limits_{i=1}^{p}\sqrt{\lambda_i\lambda_{n-i+1}}}\right)^2.$$

定理 6.6.5 在定理 6.6.3 条件下, 有

$$\mathrm{tr}X'AXX'A^{-1}X\leqslant\sum_{i=1}^{p}\frac{(\lambda_i+\lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}.$$

定理 6.6.6 设 A 为 n 阶正定 Hermite 阵, X 为 $n\times l$ 矩阵, 满足 $X^*X=I_l$, 则对任意的实数 $p>1$ 或 $p<0$, 有

$$\mathrm{tr}(X^*A^pX)\leqslant\gamma_1\cdot\mathrm{tr}(X^*AX)^p,\tag{6.6.22}$$

这里

$$\gamma_1 = \frac{(p-1)^{p-1}}{p^p} \frac{(\lambda_1^p - \lambda_n^p)^p}{(\lambda_1 - \lambda_n)(\lambda_n \lambda_1^p - \lambda_1 \lambda_n^p)^{p-1}}.$$

定理 6.6.7 在定理 6.6.6 的条件下, 对任意的实数 $p(p > 1$ 或 $p < 0)$, 有

$$\operatorname{tr}(X^* A^p X) - \operatorname{tr}(X^* AX)^p \leqslant \gamma_2 \cdot l, \tag{6.6.23}$$

这里

$$\gamma_2 = \lambda_n^p - \left(\frac{1}{p}\frac{\lambda_1^p - \lambda_n^p}{\lambda_1 - \lambda_n}\right)^{p/(p-1)} + \frac{\lambda_1^p - \lambda_n^p}{\lambda_1 - \lambda_n}\left[\left(\frac{1}{p}\frac{\lambda_1^p - \lambda_n^p}{\lambda_1 - \lambda_n}\right)^{1/(p-1)} - \lambda_n\right],$$

且对任意的实数 $0 < p < 1$, 不等式 (6.6.23) 的不等号取反向.

以上两定理可分别从推论 7.8.1 和推论 7.8.2 直接导出.

令 $p = -1$, 则

$$\frac{\operatorname{tr}(X^* AX)}{\operatorname{tr}(X^* A^{-1} X)^{-1}} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}, \tag{6.6.24}$$

$$\operatorname{tr}(X^* AX - (X^* A^{-1} X)^{-1}) \leqslant n(\sqrt{\lambda_1} - \sqrt{\lambda_n})^2.$$

注 2 由推论 6.2.3 知, $\operatorname{tr}(X^* AX)^p \leqslant (\operatorname{tr}(X^* AX))^p$, 因此, (6.6.22) 和 (6.6.23) 中 $\operatorname{tr}(X^* AX)^p$ 替换为 $(\operatorname{tr}(X^* AX))^p$ 时, 不等式仍然成立.

此外, 条件 $X^* X = I_l$ 可被替换为 $\operatorname{tr}(X^* X) = 1$, 见如下定理.

定理 6.6.8 设 A 为 n 阶正定 Hermite 阵, X 为 $n \times l$ 矩阵, 只要 $\operatorname{tr}(X^* X) = 1$, 则对任意整数 $p \neq 0, 1$, 有

$$\frac{\operatorname{tr}(X^* A^p X)}{(\operatorname{tr}(X^* AX))^p} \leqslant \frac{(p-1)^{p-1}}{p^p} \frac{(\lambda_1^p - \lambda_n^p)^p}{(\lambda_1 - \lambda_n)(\lambda_n \lambda_1^p - \lambda_1 \lambda_n^p)^{p-1}}. \tag{6.6.25}$$

与定理 6.6.3 和定理 6.6.4 相比, 显然, (6.6.24) 和 (6.6.25) 提供了更大的上界. 但当 A 的维数较高时, 计算 A 的每个特征值变得复杂. 许多软件可找到 A 的最大和最小特征值, 因此, 当 A 的维数较高时, (6.6.24) 和 (6.6.25) 会更有用.

记 $X = (x_1, \cdots, x_l)$, 于是 $\operatorname{tr}(X^* X) = 1$ 等价于 $\sum_{j=1}^{k} \|x_j\|^2 = 1$, (6.6.25) 可看作是 Ky Fan 不等式

$$\frac{\sum_{j=1}^{k} x_j^* A^p x_j}{\left(\sum_{j=1}^{k} x_j^* A x_j\right)^p} \leqslant \frac{(p-1)^{p-1}}{p^p} \frac{(\lambda_1^p - \lambda_n^p)^p}{(\lambda_1 - \lambda_n)(\lambda_n \lambda_1^p - \lambda_1 \lambda_n^p)^{p-1}}.$$

证明见 Mond 和 Pečarić(1995a).

为了便于查阅, 我们把在本书其他地方证明过的两个带约束的迹的不等式抄录在下面.

定理 6.6.9　设 A 为 $n\times n$ 正定 Hermite 阵, 则对一切满足条件 $\det X=1$ 的正定 Hermite 阵, 有

$$\mathrm{tr}AX \geqslant n(\det\mathrm{A})^{1/n}.$$

证明见定理 3.3.2.

定理 6.6.10　设 A 为 $n\times n$ 半正定 Hermite 阵, $p>1$, $\dfrac{1}{p}+\dfrac{1}{q}=1$. 则对一切满足条件 $\mathrm{tr}X^p=1$ 的半正定 Hermite 阵, 有

$$\mathrm{tr}AX \leqslant (\mathrm{tr}A^p)^{1/p},$$

等号成立 $\Longleftrightarrow X^q=\dfrac{A^p}{\mathrm{tr}A^p}$.

此即引理 6.7.3.

§6.7　矩阵的 Hölder 和 Minkowski 不等式

在经典的不等式中, 有两个著名的不等式, 一个叫 Hölder 不等式, 另一个叫 Minkowski 不等式. 它们分别有多种形式, 下面是其中的一种.

Hölder 不等式：设 $p>1$, $\dfrac{1}{p}+\dfrac{1}{q}=1$, $a_i\geqslant 0$, $b_i\geqslant 0$, $i=1,\cdots,n$. 则

$$\sum_{i=1}^{n} a_i^{1/p} b_i^{1/q} \leqslant \left(\sum_{i=1}^{n} a_i\right)^{1/p}\left(\sum_{i=1}^{n} b_i\right)^{1/q}, \tag{6.7.1}$$

且等号成立 $\Longleftrightarrow$ 存在 $c>0$, 使得 $a_i=cb_i, i=1,\cdots,n$.

Minkowski 不等式：设 $p>1$, $a_i\geqslant 0$, $b_i\geqslant 0$, $i=1,\cdots,n$. 则

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} a_i^p\right)^{1/p}+\left(\sum_{i=1}^{n} b_i^p\right)^{1/p}, \tag{6.7.2}$$

且等号成立 $\Longleftrightarrow$ 存在 $c>0$, 使得 $a_i=cb_i, i=1,\cdots,n$.

本节的目的是导出这两个不等式的矩阵推广形式, 为此我们先证明如下几个引理.

引理 6.7.1　设 $A=(a_{ij})$ 为 $n\times n$ 半正定 Hermite 阵. 则对 $p>1$, 有

$$\mathrm{tr}A^p \geqslant \sum_{i=1}^{n} a_{ii}^p.$$

当 $0<p<1$ 时, 不等号反向. 等号成立 $\Longleftrightarrow A$ 为对角阵.

证明 当 $p>1$ 时, 函数 $g(x)=x^p(x\geqslant 0)$ 为严格凸函数, 应用推论 8.3.1 知, $\sum_{i=1}^n x_i^p$ 为严格 Schur 凸函数. 于是

$$\operatorname{tr}A^p=\sum_{i=1}^n \lambda_i^p(A)=\sum_{i=1}^n g(\lambda_i(A))\geqslant\sum_{i=1}^n g(a_{ii})=\sum_{i=1}^n a_{ii}^p.$$

等号成立 $\Longleftrightarrow a=(a_{11},\cdots,a_{nn})'$ 是 $\lambda=(\lambda_1(A),\cdots,\lambda_n(A))'$ 的一个置换 $\Longleftrightarrow A$ 是对角阵. 对 $0<p<1$, 证明类似. 证毕.

下一个引理是 Hölder 不等式 (6.7.1) 的另一种叙述.

引理 6.7.2 设 $p>1, \dfrac{1}{p}+\dfrac{1}{q}=1, a_i\geqslant 0, i=1,\cdots,n$. 则对一切满足 $\sum_{i=1}^n x_i^q=1$, $x_i\geqslant 0$ 的 $x=(x_1,\cdots,x_n)'$, 有

$$a'x\leqslant\left(\sum_{i=1}^n a_i^p\right)^{1/p}, \tag{6.7.3}$$

等号成立 $\Longleftrightarrow a_1=\cdots=a_n=0$ 或 $x_i^q=\dfrac{a_i^p}{\sum_{j=1}^n a_j^p}$.

不等式 (6.7.3) 可以改写成如下极值形式

$$\max_R a'x=\left(\sum_{i=1}^n a_i^p\right)^{1/p}, \tag{6.7.4}$$

其中

$$R=\left\{x\in R^n:\ x_i\geqslant 0,\quad \sum_{i=1}^n x_i^q=1,\quad i=1,\cdots,n\right\}.$$

这个表达式是把一个非线性函数 $\left(\sum a_i^p\right)^{1/p}$ 表示成线性函数 $a'x=\sum_{i=1}^n a_ix_i$ 的包络 (envelope). 这个方法叫做拟线性化(quasilinearization), 它在不等式理论中有许多应用.

引理 6.7.3 设 $p>1, \dfrac{1}{p}+\dfrac{1}{q}=1$, A 为 $n\times n$ 半正定 Hermite 阵. 则对一切满足 $\operatorname{tr}X^q=1$ 的半正定 Hermite 阵 X, 有

$$\operatorname{tr}AX\leqslant(\operatorname{tr}A^p)^{1/p}. \tag{6.7.5}$$

等号成立 $\Longleftrightarrow X^q=\dfrac{A^p}{\operatorname{tr}A^p}$. (6.7.6)

证明　应用定理 1.4.1, 对半正定阵 Hermite 阵 X, 存在酉阵 U, 使得 $U^*XU = \Lambda = \text{diag}\ (\lambda_1, \cdots, \lambda_n)$. 定义 $B = (b_{ij}) = U^*AU$, 则

$$\text{tr}AX = \text{tr}BA = \sum_{i=1}^{n} b_{ii}\lambda_i,$$

$$\text{tr}X^q = \sum_{i=1}^{n} \lambda_i^q = 1.$$

利用引理 6.7.2, 得

$$\text{tr}AX = \sum_{i=1}^{n} b_{ii}\lambda_i \leqslant \left(\sum_{i=1}^{n} b_{ii}^p\right)^{1/p}. \tag{6.7.7}$$

因 B 也是半正定 Hermite 阵, 故应用引理 6.7.1, 我们有

$$\sum_{i=1}^{n} b_{ii}^p \leqslant \text{tr}B^p. \tag{6.7.8}$$

综合 (6.7.7) 和 (6.7.8), 得

$$\text{tr}AX \leqslant (\text{tr}B^p)^{1/p} = (\text{tr}A^p)^{1/p}.$$

在 (6.7.7) 中, 等号成立 $\Longleftrightarrow \lambda_i = \dfrac{b_{ii}^p}{\sum\limits_{j=1}^{n} b_{ii}^p}$, $i = 1, \cdots, n$. 而在 (6.7.8) 中, 等号成立 $\Longleftrightarrow B$ 为对角阵. 于是 (6.7.5) 等号成立 $\Longleftrightarrow A^q = \dfrac{B^p}{\text{tr}B^p}$, 这又等价于 (6.7.6). 证毕.

注 1　引理 6.7.3 能够看作引理 6.7.2 的一种推广. 我们也可以把它改写成极值形式.

记

$$R = \{X\text{：}\ X \text{ 为 } n \times n \text{ 半正定 Hermite 阵},\ \text{tr}X^q = 1\},$$

则引理 6.7.3 可叙述为：对任意 $n \times n$ 半正定 Hermite 阵 A, 有

$$\max_{X \in R} \text{tr}AX = (\text{tr}A^p)^{1/p}. \tag{6.7.9}$$

这里和 (6.7.4) 式相类似, 我们把 A 的非线性函数 $(\text{tr}A^p)^{1/p}$ 表示为 A 的线性函数 $\text{tr}AX$ 的包络. 换句话说, 我们获得了 $(\text{tr}A^p)^{1/p}$ 的拟线性化表示.

在做了上面的准备之后, 我们可以很容易获得矩阵的 Hölder 不等式和 Minkowski 不等式. 它们是由 Magnus(1987) 证明的.

定理 6.7.1(矩阵的 Hölder 不等式)　设 A 和 B 为两个同阶半正定 Hermite 阵, $A \neq 0$, $B \neq 0$, $p > 1$, $\dfrac{1}{p} + \dfrac{1}{q} = 1$. 则

$$\text{tr}A^{1/p}B^{1/q} \leqslant (\text{tr}A)^{1/p}(\text{tr}B)^{1/q}, \tag{6.7.10}$$

等号成立 $\Longleftrightarrow$ 存在数 $c>0$, 使得 $B=cA$.

证明 定义

$$X=\frac{B^{1/q}}{(\mathrm{tr}B)^{1/q}},$$

显然, 这样的 X 满足 $\mathrm{tr}X^q=1$, 且是半正定 Hermite 阵. 应用引理 6.7.3, 用 $A^{1/p}$ 代替 A, 便得到 (6.7.10). 并且等号成立 $\Longleftrightarrow \dfrac{B}{\mathrm{tr}B}=\dfrac{A}{\mathrm{tr}A} \Longleftrightarrow$ 存在 $c>0$, 使得 $B=cA$. 定理证毕.

定理 6.7.2(矩阵的 Minkowski 不等式) 对任意两个同阶半正定 Hermite 阵 A 和 B, $A\neq 0$, $B\neq 0$, $p<1$, 有

$$(\mathrm{tr}(A+B)^p)^{1/p}\leqslant(\mathrm{tr}A^p)^{1/p}+(\mathrm{tr}B^p)^{1/p}. \tag{6.7.11}$$

等号成立 $\Longleftrightarrow$ 存在 $c>0$, 使得 $A=cB$.

证明 记 $q=\dfrac{p}{p-1}$, $R=\{X$: X 为与 A 同阶的半正定 Hermite 阵, $\mathrm{tr}X^q=1\}$. 由拟线性化表示 (6.7.9), 得

$$\begin{aligned}(\mathrm{tr}(A+B)^p)^{1/p}&=\max_R \mathrm{tr}(A+B)X\\&\leqslant \max_R \mathrm{tr}AX+\max_R \mathrm{tr}BX\\&=(\mathrm{tr}A^p)^{1/p}+(\mathrm{tr}B^p)^{1/p}.\end{aligned} \tag{6.7.12}$$

上式等号成立 $\Longleftrightarrow$ 存在 X 能够使 $\mathrm{tr}AX$, $\mathrm{tr}BX$, $\mathrm{tr}(A+B)X$ 同时达到最大 $\Longleftrightarrow \dfrac{A^p}{\mathrm{tr}A^p}=\dfrac{B^p}{\mathrm{tr}B^p}=\dfrac{(A+B)^p}{\mathrm{tr}(A+B)^p} \Longleftrightarrow$ 存在 $c>0$, 使得 $A=cB$. 证毕.

§6.8 其 他

本节讨论一些迹的不等式, 它们很难分类到前面任何一节.

定理 6.8.1 设 A 为 n 阶正定 Hermite 阵, B 为 n 阶半正定 Hermite 阵, 且 $A\geqslant B$. 则

$$\frac{\mathrm{tr}B}{\mathrm{tr}A}\geqslant\frac{\det B}{\det A}.$$

证明 显然, 当 $\det B=0$ 时, 不等式成立. 故以下假设 $B>0$. 记 $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 和 $\mu_1\geqslant\cdots\geqslant\mu_n$ 分别为 A 和 B 的特征值. 由推论 4.4.1 知, $\lambda_i\geqslant\mu_i$, $i=1,\cdots,n$. 于是问题归结为对满足这些条件的两组正数, 证明

$$\frac{\sum\limits_{i=1}^n \mu_i}{\sum\limits_{i=1}^n \lambda_i}\geqslant\frac{\prod\limits_{i=1}^n \mu_i}{\prod\limits_{i=1}^n \lambda_i}. \tag{6.8.1}$$

我们对阶数 n 施归纳法. 当 $n=2$ 时, 我们需证明

$$\frac{\mu_1+\mu_2}{\lambda_1+\lambda_2} \geqslant \frac{\mu_1\mu_2}{\lambda_1\lambda_2}, \tag{6.8.2}$$

它等价于

$$\lambda_1\mu_1(\lambda_2-\mu_2)+\lambda_2\mu_2(\lambda_1-\mu_1) \geqslant 0.$$

因为 $\lambda_i \geqslant \mu_i$, $i=1,2$, 于是 (6.8.2) 得证.

假设对 $n-1$ 命题成立, 则利用 (6.8.2), 对一般的 n, 有

$$\frac{(\mu_1+\mu_2)+\cdots+\mu_n}{(\lambda_1+\lambda_2)+\cdots+\lambda_n} \geqslant \frac{(\mu_1+\mu_2)\cdot\mu_3\cdots\mu_n}{(\lambda_1+\lambda_2)\cdot\lambda_3\cdots\lambda_n} \geqslant \frac{\mu_1\mu_2\cdots\mu_n}{\lambda_1\lambda_2\cdots\lambda_n}.$$

证毕.

定理 6.8.2　设 A 为 $n\times n$ 正定 Hermite 阵, B 和 C 为 $n\times n$ 半正定 Hermite 阵, 且 $B\geqslant C$. 则

$$\operatorname{tr}(A+B)^{-1}B \geqslant \operatorname{tr}(A+C)^{-1}C.$$

证明　首先

$$\begin{aligned}\operatorname{tr}(A+B)^{-1}B &= \operatorname{tr}[(A+B)^{-1}((A+B)-A)] \\ &= n-\operatorname{tr}(A+B)^{-1}A, \end{aligned} \tag{6.8.3}$$

$$\operatorname{tr}(A+C)^{-1}C = n-\operatorname{tr}(A+C)^{-1}C. \tag{6.8.4}$$

因为 $B\geqslant C$, 故 $A+B\geqslant A+C$, 应用推论 7.2.2, 得

$$(A+C)^{-1} \geqslant (A+B)^{-1}.$$

因而

$$A^{1/2}(A+B)^{-1}A^{1/2} \leqslant A^{1/2}(A+C)^{-1}A^{1/2}.$$

两边取迹, 我们有

$$\operatorname{tr}(A+B)^{-1}A \leqslant \operatorname{tr}(A+C)^{-1}A.$$

再利用 (6.8.3) 和 (6.8.4), 定理得证.

最后我们证明一个由某一矩阵的迹为元素的矩阵的半正定性.

定理 6.8.3　设 A 为半正定 Hermite 阵, 且分块为

$$A=\begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mm} \end{pmatrix},$$

其中 A_{ij} 为 $n\times n$ 阵, 则 $m\times m$ 的方阵

$$T=(\mathrm{tr}(A_{ij}))\geqslant 0.$$

证明 记 $e_i=(0,\cdots,0,1,0,\cdots,0)'$, 即第 i 个元素为 1, 其余所有元素为零的 n 维向量. 则对任意 n 阶方阵 B

$$\mathrm{tr}B=\sum_{i=1}^{n}e_i'Be_i,$$

于是

$$\begin{aligned}T=&\sum_{i=1}^{n}\begin{pmatrix}e_i'A_{11}e_i & \cdots & e_i'A_{1m}e_i\\ \vdots & & \vdots\\ e_i'A_{m1}e_i & \cdots & e_i'A_{mm}e_i\end{pmatrix}\\ =&\sum_{i=1}^{n}\begin{pmatrix}e_i' & 0 & \cdots & 0\\ 0 & e_i' & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & e_i'\end{pmatrix}\begin{pmatrix}A_{11} & \cdots & A_{1m}\\ \vdots & & \vdots\\ A_{m1} & \cdots & A_{mm}\end{pmatrix}\begin{pmatrix}e_i & 0 & \cdots & 0\\ 0 & e_i & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & e_i\end{pmatrix}.\end{aligned}$$

定理得证.

第7章 偏　　序

§7.1 定　　义

所谓偏序 (partial ordering), 乃是指在一个集合 S 上定义的一种关系, 记为“$\succ$”, 它满足如下三条:

(1) 自反性: $x \succ x$, 对一切 $x \in S$ 成立;

(2) 传递性: 若 $x \succ y$, $y \succ z$, 则 $x \succ z$;

(3) 若 $x \succ y$, $y \succ x$, 则 $x = y$.

如果 x 和 y 皆为 S 中元素, 且 $x \succ y$ 或 $y \succ x$ 至少一个成立, 则称 x 与 y 是可比的. 在定义了偏序的集合中, 并不是任何两个元素都是可比的, 这也就是称其为偏序的原因.

设 S 为 n 阶方阵的全体, 若 $B - A \geqslant 0$, 即 $B - A$ 为半正定阵, 则称 A 低于 B, 记为 $B \geqslant A$ 或 $A \leqslant B$. 容易验证, S 上的这种关系是一种偏序, 称为 Löwner 偏序. 特别, 当 $B - A > 0$ 时, 记为 $B > A$ 或 $A < B$.

在矩阵论文献中, 还存在另外两种偏序. 其一是秩减 (rank subtractivity) 偏序, 其定义为: 若 $r(B-A) = r(B) - r(A)$, 则称 A 在秩减意义下低于 B. 这种偏序关系是由 Hartwig(1980) 首先讨论的. 文献中有时也称减号 (minus) 偏序 (见 Baksalary (1986) 和 Mitra(1986)). Baksalary 和 Hauke(1984) 证明了: 对于半正定实对称阵, 秩减偏序关系蕴含着 Löwner 偏序. 即, 若 A 和 B 为两个半正定实对称阵, 如果在秩减意义下 A 低于 B, 则在 Löwner 意义下, A 也低于 B.

另一种偏序是由 Drazin(1978) 引进的, 称为 Drazin 偏序或星 (star) 偏序, 其定义是: 若 $A^*A = A^*B$, $AA^* = BA^*$, 则称 A 在 Drazin 或星意义下低于 B. 可以证明 (见 Hartwig 和 Styan(1987)) 对任意两个矩阵, 由星偏序可以推出秩减偏序.

本章只讨论 Löwner 偏序. 在 §7.2, 我们将叙述 $A \geqslant B$ 和 $A > B$ 的基本性质及成立的充要条件; §7.3 则研究 $A^2 \geqslant B^2$ 的类似问题及其与 $A \geqslant B$ 的关系, 并提及 $A^k \geqslant B^k$ 的一些结果; §7.4 给出有关正定阵的主子阵的一个偏序关系; §7.5~§7.8 分别把 Cauchy-Schwarz 不等式、Kantorovich 不等式、Wielandt 不等式以及凸函数不等式推广到矩阵形式; 最后一节研究了矩阵 Hadamard 乘积的偏序.

§7.2 $A \geqslant B$

假定 A 和 B 为两个 n 阶 Hermite 阵, 本节研究 $A \geqslant B$ 和 $A > B$ 的基本性质

及这些偏序成立的充要条件.

我们先把推论 4.4.1 转述在这里, 并给出一个推论.

定理 7.2.1 设 A 和 B 为两个 n 阶 Hermite 阵.

(1) 若 $A \geqslant B$, 则 $\lambda_i(A) \geqslant \lambda_i(B), i=1,\cdots,n$;

(2) 若 $A > B$, 则 $\lambda_i(A) > \lambda_i(B), i=1,\cdots,n$.

由此立即可得如下推论.

推论 7.2.1 设 $A \geqslant B \geqslant 0$. 则

(1) $\mathrm{tr}A \geqslant \mathrm{tr}B$;

(2) $\det A \geqslant \det B$;

(3) $r(A) \geqslant r(B)$.

若 $A > B \geqslant 0$, 则在 (1) 和 (2) 中严格不等号成立.

定理 7.2.2 设 A 和 B 为两个 n 阶 Hermite 阵, P 为 $n \times k$ 矩阵.

(1) 若 $A \geqslant B$, 则 $P^*AP \geqslant P^*BP$;

(2) 若 $r(P)=k$, $A > B$, 则 $P^*AP > P^*BP$.

证明 (1) 由 $A \geqslant B$ 之定义知, 对任意 $x \in R^n$, 有 $x^*(A-B)x \geqslant 0$. 于是对任意 $x \in R^n$,

$$x^*(P^*AP - P^*BP)x = (Px)^*(A-B)(Px) \geqslant 0,$$

此即 $P^*AP - P^*BP \geqslant 0$. 定理证毕.

(2) 的证明与 (1) 相类似, 留给读者做练习.

定理 7.2.3 设 A 和 B 为两个 n 阶 Hermite 阵. $A > 0$, $B \geqslant 0$, 则

(1) $A \geqslant B \Longleftrightarrow \lambda_1(BA^{-1}) \leqslant 1$;

(2) $A > B \Longleftrightarrow \lambda_1(BA^{-1}) < 1$.

这里 $\lambda_1(A)$ 表示 A 的最大特征值.

证明 根据定理 7.2.2, 我们有

$$\begin{aligned}
A \geqslant B &\Longleftrightarrow A^{-1/2}(A-B)A^{-1/2} \geqslant 0 \\
&\Longleftrightarrow I_n - A^{-1/2}BA^{-1/2} \geqslant 0 \\
&\Longleftrightarrow \lambda_i(I_n - A^{-1/2}BA^{-1/2}) \geqslant 0,\ i=1,\cdots,n(\text{ 由定理 } 1.4.2(1)\,) \\
&\Longleftrightarrow \lambda_1(A^{-1/2}BA^{-1/2}) \leqslant 1 \\
&\Longleftrightarrow \lambda_1(BA^{-1}) \leqslant 1,
\end{aligned}$$

其中 $\lambda_i(A)$ 表示方阵 A 的第 i 个顺序特征值.

(1) 得证, 同法可证 (2), 定理证毕.

应用定理 7.2.3, 立即可得如下推论.

推论 7.2.2 设 A 和 B 皆为正定 Hermite 阵. 则

(1) $A \geqslant B \Longleftrightarrow B^{-1} \geqslant A^{-1}$;

(2) $A > B \Longleftrightarrow B^{-1} > A^{-1}$.

下面的定理将对 $A \geqslant B$ 或 $A > B$ 作进一步表征. 为此我们需要如下引理.

引理 7.2.1　设 $A \geqslant B \geqslant 0$, 则 $\mathscr{M}(B) \subset \mathscr{M}(A)$, 这里 $\mathscr{M}(A)$ 表示矩阵 A 的列向量张成的子空间.

证明　首先从定义知: $A \geqslant B \Longleftrightarrow$ 对任意 x, $x^*Ax \geqslant x^*Bx$. 若 $x \in \mathscr{M}(A)^{\perp}$, 也就是说, x 为 $\mathscr{M}(A)$ 的正交补空间上的任意向量, 则 $x^*Ax = 0$, 进而有 $x^*Bx = 0$, 也就是 $x \in \mathscr{M}(B)^{\perp}$. 这就证明了

$$\mathscr{M}(A)^{\perp} \subset \mathscr{M}(B)^{\perp}.$$

应用定理 1.1.3(3), 得 $\mathscr{M}(B) \subset \mathscr{M}(A)$. 证毕.

引理 7.2.2　设 A 和 B 为半正定 Hermite 阵, 则 $A \geqslant B$ 当且仅当 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 和 $A(A-B)A \geqslant 0$.

证明　依半正定的定义及引理 7.2.1, 必要性是显然的. 下证充分性. 易见我们的问题归结为: 在条件 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 下, 证明

$$A \geqslant B \Longleftrightarrow A(A-B)A \geqslant 0. \tag{7.2.1}$$

因为 AA^+ 是向 $\mathscr{M}(A)$ 的正交投影阵, 则由 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 可推得

$$AA^+(A-B)AA^+ = A - B.$$

于是, 在条件 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 下,

$$A \geqslant B \Longleftrightarrow x^*AA^+(A-B)AA^+x \geqslant 0, \quad \text{对任意 } x \in R^n.$$

注意到, $\mathscr{M}(AA^+) = \mathscr{M}(A)$, 因而,

$$x^*AA^+(A-B)AA^+x \geqslant 0, \text{对任意 } x \in R^n$$

等价于

$$x^*A(A-B)Ax \geqslant 0, \quad \text{对任意 } x \in C^n,$$

(7.2.1) 得证. 引理证毕.

定理 7.2.4　设 A 和 B 为 n 阶 Hermite 阵, 且 $A \geqslant 0$, $B \geqslant 0$. 则 $A \geqslant B$ 当且仅当 $\mathscr{M}(B) \subset \mathscr{M}(A)$, $\lambda_1(BA^-) \leqslant 1$, 这里最大特征值 $\lambda_1(BA^-)$ 与 A^- 选择无关.

证明　根据引理 7.2.1. 问题归结为, 在条件 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 下, 证明 $A \geqslant B$ 与 $\lambda_1(BA^-) \leqslant 1$ 等价, 再结合引理 7.2.2, 问题进一步化为, 在条件 $\mathscr{M}(B) \subset \mathscr{M}(A)$ 条件下, 证明 $A(A-B) \times A \geqslant 0 \Longleftrightarrow \lambda_1(BA^-) \leqslant 1$.

现对 A 作满秩分解 (参考定理 1.5.3): $A = LL^*$, 这里 L 为 $n \times k$ 矩阵, $k = r(A)$. 则

$$A(A-B)A \geqslant 0 \Longleftrightarrow L^*(A-B)L \geqslant 0.$$

用 $(L^*L)^{-1}$ 去右乘和左乘上式, 知

$$\begin{aligned} \text{上式} &\Longleftrightarrow I - T^*BT \geqslant 0, \quad \text{这里 } T = L(L^*L)^{-1} \\ &\Longleftrightarrow \lambda_1(T^*BT) \leqslant 1 \\ &\Longleftrightarrow \lambda_1(BTT^*) \leqslant 1 \\ &\Longleftrightarrow \lambda_1(BA^+) \leqslant 1. \end{aligned}$$

在最后一步, 我们应用了 $TT^* = L(L^*L)^{-2}L^* = A^+$.

最后我们尚需证明, 对任一广义逆 A^-, $\lambda_1(BA^-) = \lambda_1(BA^+)$. 事实上, 因为 $\mathscr{M}(B) \subset \mathscr{M}(A)$, 因而 $B^{1/2}A^-B^{1/2}$ 与广义逆 A^- 的选择无关, 于是 $B^{1/2}A^-B^{1/2} = B^{1/2}A^+B^{1/2}$, 对一切 A^- 成立. 故有

$$\lambda_1(BA^-) = \lambda_1(B^{1/2}A^-B^{1/2}) = \lambda_1(B^{1/2}A^+B^{1/2}) = \lambda_1(BA^+).$$

定理证毕.

注 1 Stepniak, Wang 和 Wu (1984) 给出 $A \geqslant B$ 的另一种表征, 即若 A 和 B 为 Hermite 阵, $A \geqslant 0$, $B \geqslant 0$, 则 $A \geqslant B \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A)$, $B \geqslant BA^-B$. 并把这个结果应用于线性模型参数估计理论 (详见 §8.1). 在他们的证明中, 应用了如下事实: 设 A 和 B 为半正定 Hermite 阵, 且 $A \geqslant B$, 则一定存在 B^- 和 A^-, 使得 $B^- \geqslant A^-$, 证明可以在 Wu(1980) 中找到. 这是推论 7.2.2 的一个推广. 但是一般说来, 由 $A \geqslant B$ 并不能推出对任意 B^- 和 A^-, 有 $B^- \geqslant A^-$. 但对 Moore-Penrose 广义逆, 当 $r(A) = r(B)$ 时, 我们有 $A \geqslant B \Longleftrightarrow B^+ \geqslant A^+$. 这就是下面的定理.

定理 7.2.5 设 A 和 B 为半正定 Hermite 阵, 且 $r(A) = r(B)$. 则

$$A \geqslant B \Longleftrightarrow B^+ \geqslant A^+.$$

证明 因为 $\mathscr{M}(B) = \mathscr{M}(B^+)$, $\mathscr{M}(A) = \mathscr{M}(A^+)$, 应用定理 7.2.4, 我们有

$$\begin{aligned} A \geqslant B &\Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A), \lambda_1(BA^+) \leqslant 1 \\ &\Longleftrightarrow \mathscr{M}(B^+) \subset \mathscr{M}(A^+), \lambda_1(BA^+) \leqslant 1, \end{aligned}$$

再结合 $r(A) = r(B)$, 知

$$\text{上式} \Longleftrightarrow \mathscr{M}(B^+) = \mathscr{M}(A^+), \ \lambda_1(BA^+) \leqslant 1.$$

注意到, $\lambda_1(BA^+) = \lambda_1(A^+(B^+)^+)$, 再次利用定理 7.2.4, 得知上式等价于 $B^+ \geqslant A^+$. 证毕.

在这个定理中, 条件 $r(A) = r(B)$ 是非常重要的. 读者也很容易举例说明, 当 $r(A) \neq r(B)$ 时, 由 $A \geqslant B$ 并不能推出 $B^+ \geqslant A^+$. 下面的定理证明了, 在 $A \geqslant B$ 条件下, $B^+ \geqslant A^+$ 与 $r(A) = r(B)$ 等价.

定理 7.2.6 设 A 和 B 皆为半正定 Hermite 阵, 且 $A \geqslant B$. 则

$$B^+ \geqslant A^+ \Longleftrightarrow r(A) = r(B).$$

证明 充分性已在定理 7.2.5 证明过. 以下证必要性. 应用引理 7.2.1, 我们有

$$A \geqslant B \Longrightarrow \mathscr{M}(B) \subset \mathscr{M}(A),$$
$$B^+ \geqslant A^+ \Longrightarrow \mathscr{M}(A^+) \subset \mathscr{M}(B^+).$$

于是

$$\mathscr{M}(B) \subset \mathscr{M}(A) = \mathscr{M}(A^+) \subset \mathscr{M}(B^+) = \mathscr{M}(B).$$

由此可得: $\mathscr{M}(A) = \mathscr{M}(B)$, 因而 $r(A) = r(B)$. 证毕.

综合定理 7.2.5 和定理 7.2.6, 我们有

推论 7.2.3 设 A 和 B 为同阶半正定 Hermite 阵. 则下列任意两条可以推出第三条.

(1) $A \geqslant B$;

(2) $B^+ \geqslant A^+$;

(3) $r(A) = r(B)$.

在前面所有结果中, 我们都要求 Hermite 阵 A 和 B 是半正定 (或正定) 的, 这一点是必要的. 例如, 在推论 7.2.2(2) 中, 若取

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} > 0,$$

但

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

显然 B 不是正定阵. 虽然

$$A - B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} > 0,$$

即 $A > B$ 成立, 但

$$B^{-1} - A^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}$$

不是正定的, 即 $B^{-1} > A^{-1}$ 不成立. 这表明, 当 A 和 B 为 Hermite 阵, 但放弃半正定或正定假设时, 需要另外的补充条件才能保证 $A \geqslant B$ 与 $B^{-1} \geqslant A^{-1}$ 的等价性. 在这方面 Liski(1992) 用很初等方法建立了一些结果. 我们先证明一个预备事实.

引理 7.2.3 设 A 和 B 均为 n 阶可逆的 Hermite 阵, 且 $A \geqslant B$. 则 $B^{-1}A$ 的特征值都是实数.

证明 由 $A \geqslant B$ 知 $A = B + KK^*$, 这里 K 为 $n \times q$ 矩阵. 由满秩分解定理 (见定理 1.5.3), 我们可以假设 $r(K) = q$, 因而 $B^{-1}A = I_n + B^{-1}KK^*$. 故

$$\lambda(B^{-1}A) = 1 + \lambda(B^{-1}KK^*), \tag{7.2.2}$$

这里 $\lambda(A)$ 表示 A 的任一特征值. 因为 $B^{-1}KK^*$ 与 $K^*B^{-1}K$ 具有相同的非零特征值, 因而从 (7.2.2) 知, $B^{-1}A$ 的特征值或为 1, 或为

$$\lambda(B^{-1}A) = 1 + \lambda(K^*B^{-1}K).$$

但 $K^*B^{-1}K$ 为 Hermite 阵, 所以 $K^*B^{-1}K$ 的特征值为实数. 于是 $B^{-1}A$ 的特征值全为实数. 证毕.

定理 7.2.7(Liski) 设 A 和 B 皆为 n 阶可逆 Hermite 阵, 且 $\lambda_n(B^{-1}A) > 0$. 则 $A \geqslant B \Longleftrightarrow B^{-1} \geqslant A^{-1}$.

证明 必要性. 和引理 7.2.3 一样, A 可表为 $A = B + KK^*$, 这里 $r(K) = q$. 所以, $B^{-1}A$ 的那些不等于 1 的特征值可表为

$$\begin{aligned} \lambda(B^{-1}A) &= \lambda(I_n + B^{-1}KK^*) = 1 + \lambda(B^{-1}KK^*) \\ &= 1 + \lambda(K^*B^{-1}K) = \lambda(I_q + K^*B^{-1}K). \end{aligned}$$

由 $B^{-1}A$ 可逆, 可推得 $I_q + K^*B^{-1}K$ 也可逆. 再由假设 $\lambda_n(B^{-1}A) > 0$, 知

$$I_q + K^*B^{-1}K > 0. \tag{7.2.3}$$

另一方面, 利用公式

$$(M + CNC^*)^{-1} = M^{-1} - M^{-1}C(C^*M^{-1}C + N)^{-1}C^*M^{-1}, \tag{7.2.4}$$

我们有

$$\begin{aligned} B^{-1} - A^{-1} &= B^{-1} - (B + KK^*)^{-1} \\ &= B^{-1}K(I_q + K^*B^{-1}K)^{-1}K^*B^{-1}, \end{aligned}$$

结合 (7.2.3), 推得 $B^{-1}-A^{-1}>0$, 即 $B^{-1}>A^{-1}$. 必要性得证.

充分性. 由 $B^{-1}\geqslant A^{-1}$, 知 $B^{-1}=A^{-1}+CC^*$, 这里 C 为 $n\times p$ 矩阵, $r(C)=p$. 于是 $B^{-1}A=I_n+CC^*A$. 再次利用 (7.2.4), 得

$$B=(A^{-1}+CC^*)^{-1}=A-AC(C^*AC+I_p)^{-1}C^*A,$$

即

$$A-B=AC(I_p+C^*AC)^{-1}C^*A. \tag{7.2.5}$$

类似于必要性部分的讨论, 由 $B^{-1}A$ 的所有特征值为正, 知 $I_p+C^*AC>0$. 结合 (7.2.5) 就证明了 $A-B>0$, 即 $A>B$. 定理证毕.

在结束这一节的时候, 我们给出 $A-b_1b_1^*\geqslant 0$ 和 $A-b_1b_1^*-b_2b_2^*\geqslant 0$ 的充要条件. 这些条件都可以从下面的定理推出.

定理 7.2.8 设 A 为 $n\times n$ Hermite 阵, D 为 $n\times p$ 矩阵. 则 $A-DD^*\geqslant 0\Longleftrightarrow A\geqslant 0$, $\mathscr{M}(D)\subset\mathscr{M}(A)$, $I_p-D^*A^-D\geqslant 0$. 且左边严格不等式成立 $\Longleftrightarrow$ 右边两个严格不等式成立.

证明 因为 $A-DD^*\geqslant 0$. 由引理 7.2.1 知 $\mathscr{M}(DD^*)\subset\mathscr{M}(A)$. 但 $\mathscr{M}(D)=\mathscr{M}(DD^*)$, 于是 $\mathscr{M}(D)\subset\mathscr{M}(A)$. 可见, 我们的问题归结为在条件 $\mathscr{M}(D)\subset\mathscr{M}(A)$ 下, 证明

$$A-DD^*\geqslant 0\Longleftrightarrow A\geqslant 0, I_p-D^*A^-D\geqslant 0. \tag{7.2.6}$$

因为由 $\mathscr{M}(D)\subset\mathscr{M}(A)$ 可推出, 存在矩阵 T, 使得 $D=AT$. 因而

$$D^*-D^*A^-A=T^*A-T^*AA^-A=0,$$
$$D-AA^-D=AT-AA^-AT=0.$$

于是

$$\begin{pmatrix} I_n & 0 \\ -D^*A^- & I_p \end{pmatrix}\begin{pmatrix} A & D \\ D^* & I_p \end{pmatrix}\begin{pmatrix} I_n & -A^-D \\ 0 & I_p \end{pmatrix}=\begin{pmatrix} A & 0 \\ 0 & I_p-D^*A^-D \end{pmatrix}. \tag{7.2.7}$$

另一方面

$$\begin{pmatrix} I_p & 0 \\ -D & I_n \end{pmatrix}\begin{pmatrix} I_p & D^* \\ D & A \end{pmatrix}\begin{pmatrix} I_p & -D^* \\ 0 & I_n \end{pmatrix}=\begin{pmatrix} I_p & 0 \\ 0 & A-DD^* \end{pmatrix}, \tag{7.2.8}$$

若记

$$M_1=\begin{pmatrix} A & D \\ D^* & I_p \end{pmatrix}, \qquad M_2=\begin{pmatrix} I_p & D^* \\ D & A \end{pmatrix},$$

显然 $M_1 \geqslant 0 \Longleftrightarrow M_2 \geqslant 0$. 但由 (7.2.7) 知 $M_1 \geqslant 0 \Longleftrightarrow A \geqslant 0$, $I_p - D^*A^-D \geqslant 0$. 再由 (7.2.8) 知 $M_2 \geqslant 0 \Longleftrightarrow A - DD^* \geqslant 0$. 这就证明了 (7.2.6). 关于严格不等式成立的结论是明显的. 定理证毕.

注 2 在定理条件下, $\mathscr{M}(D) \subset \mathscr{M}(A)$ 总是成立, 于是 D^*A^-D 与广义逆 A^- 的选择无关.

在定理中, 若 $p=1$, 我们便得到

推论 7.2.4 设 A 为 $n \times n$ Hermite 阵, b 为 $n \times 1$ 向量. 则

$$A - bb^* \geqslant 0 \Longleftrightarrow A \geqslant 0, \quad b \in \mathscr{M}(A), \quad b^*A^-b \leqslant 1.$$

进一步,

$$A - bb^* > 0 \Longleftrightarrow A > 0, \quad b^*A^{-1}b < 1.$$

下面的推论是定理中 $p=2$ 的特殊情形.

推论 7.2.5 设 A 为 $n \times n$ Hermite 阵. b_1 和 b_2 为线性无关向量. 记 $c_{ij} = b_i^*A^-b_j, i = 1,2$. 则

$$A - b_1b_1^* - b_2b_2^* \geqslant 0 \tag{7.2.9}$$

当且仅当

$$A \geqslant 0, \tag{7.2.10}$$

$$b_i \in \mathscr{M}(A), \quad i = 1,2, \tag{7.2.11}$$

$$c_{ii} \leqslant 1, \quad i = 1,2, \tag{7.2.12}$$

$$|c_{12}|^2 \leqslant (1 - c_{11})(1 - c_{22}), \tag{7.2.13}$$

且 (7.2.9) 中严格不等式成立 $\Longleftrightarrow$ (7.2.10), (7.2.12) 和 (7.2.13) 中严格不等式成立.

证明 在定理 7.2.8 中, 取 $D = (b_1, b_2)$, 知 (7.2.9) 等价于 (7.2.10), (7.2.11) 和

$$\begin{pmatrix} 1 - c_{11} & -c_{12} \\ -c_{21} & 1 - c_{12} \end{pmatrix} \geqslant 0. \tag{7.2.14}$$

但 (7.2.14) 又等价于 (7.2.12) 和 (7.2.13). 其余结论的证明是容易的. 证毕.

关于 $A - b_1b_1^* - b_2b_2^* \geqslant 0 (> 0)$ Baksalary 等 (1990) 还建立了其他一些充要条件. 读者从推论 7.2.5 不难看到, 应用定理 7.2.8, 我们还可以建立矩阵

$$C(k) = A - \sum_{i=1}^{k} b_ib_i^*, \ k \geqslant 3$$

的半正定 (正定) 的充要条件.

§7.3　$A^2 \geqslant B^2$

对于两个非负数 a 和 b, $a \geqslant b$ 与 $a^2 \geqslant b^2$ 等价. 但是对于两个半正定 Hermite 阵, 类似的事实并不成立, 即 $A \geqslant B$ 与 $A^2 \geqslant B^2$ 并不等价. 本节我们先给出 $A^2 \geqslant B^2$ 的一个表征, 尔后证明由 $A^2 \geqslant B^2$ 能够推出 $A \geqslant B$, 但反过来, 一般是不对的. 然而若 AB 可交换, 即 $AB = BA$, 则由 $A \geqslant B$ 推出 $A^2 \geqslant B^2$.

定理 7.3.1　设 A 和 B 为两个 Hermite 阵. 则 $A^2 \geqslant B^2 \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A)$, 且 $\sigma_1(BA^+) \leqslant 1$. 其中 $\sigma_1(BA^+)$ 表示 BA^+ 的最大奇异值.

证明　应用定理 7.2.4, 我们有

$$A^2 \geqslant B^2 \Longleftrightarrow \mathscr{M}(B^2) \subset \mathscr{M}(A^2), \quad \lambda_1(B^2(A^2)^+) \leqslant 1.$$

这里 $\lambda_1(A)$ 表示 A 的最大特征值. 利用 $\mathscr{M}(A) = \mathscr{M}(A^2)$, $\mathscr{M}(B) = \mathscr{M}(B^2)$ 以及

$$\begin{aligned}\lambda_1^{1/2}(B^2(A^2)^+) &= \lambda_1^{1/2}(A^+B^2A^+) = \lambda_1^{1/2}((BA^+)^*(BA^+))\\ &= \sigma_1(BA^+),\end{aligned}$$

定理得证.

为了证明下面一个定理, 我们先证如下引理.

引理 7.3.1　设 A 为复方阵, $\lambda_1(A)$ 和 $\sigma_1(A)$ 分别为它的最大特征值和最大奇异值. 则 $|\lambda_1(A)| \leqslant \sigma_1(A)$.

证明　设 x 为 A 的对应于 $\lambda_1(A)$ 的单位特征向量, 则

$$|\lambda_1(A)|^2 = \lambda_1\overline{\lambda}_1 = x^*A^*Ax \leqslant \lambda_1(A^*A) = \sigma_1^2(A),$$

上式不等号处利用了 Rayleigh-Ritz 定理 (见定理 4.1.1). 定理证毕.

定理 7.3.2　设 A 和 B 为两个半正定 Hermite 阵. 则

(1) $A^2 \geqslant B^2 \Longrightarrow A \geqslant B$;

(2) 若 $AB = BA$, 则 $A \geqslant B \Longrightarrow A^2 \geqslant B^2$.

证明　(1) 应用定理 7.3.1 知

$$A^2 \geqslant B^2 \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A), \sigma_1(BA^+) \leqslant 1. \tag{7.3.1}$$

而由定理 7.2.4 可得

$$A \geqslant B \Longleftrightarrow \mathscr{M}(B) \subset \mathscr{M}(A), \lambda_1(BA^+) \leqslant 1. \tag{7.3.2}$$

注意到, 当 A 和 B 为半正定 Hermite 阵时, $\lambda_1(BA^+)$ 为正数, 故由引理 7.3.1, 有

$$\lambda_1(BA^+) \leqslant \sigma_1(BA^+).$$

这样, 由 (7.3.1) 和 (7.3.2) 知, 对半正定 Hermite 阵 A 和 B, 从 $A^2 \geqslant B^2$ 可推得 $A \geqslant B$.

(2) 比较 (7.3.1) 和 (7.3.2) 容易明白, 我们只需证明：在条件 $AB = BA$ 下.

$$\lambda_1(BA^+) \leqslant 1 \Longrightarrow \sigma_1(BA^+) \leqslant 1. \tag{7.3.3}$$

因为 $AB = BA$, 故 A 和 B 可以同时对角化. 即存在酉阵 Q, 使得

$$A = Q\Lambda Q^*, \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),\ \lambda_i \geqslant 0,\ i = 1, \cdots, n,$$

$$B = Q\Delta Q^*, \Delta = \mathrm{diag}(\delta_1, \cdots, \delta_n),\ \delta_i \geqslant 0,\ i = 1, \cdots, n.$$

据此容易证得 $A^+B = BA^+$. 于是 BA^+ 为 Hermite 阵, 且是半正定的. 故

$$\sigma_1(BA^+) = \lambda_1(BA^+) \leqslant 1,$$

这就证明了 (7.3.3). 定理证毕.

应用这个定理可以推出矩阵偏序的其他一些结果, 详见 Baksalary, Schipp 和 Trenkler(1992).

最后作为定理 7.3.2 的一个推广, 我们不加证明地叙述关于 Hermite 阵幂的乘积的偏序的一个更一般结果, 它是由 Chan 和 Kwong 于 1985 年建立的.

定理 7.3.3 设 A, B, C 和 D 为同阶半正定 Hermite 阵, 且 A 与 C 可交换, B 与 D 也可交换, $A \geqslant B \geqslant 0, C \geqslant D \geqslant 0$. 则对任意 $r + s \leqslant 1$ 的正数 r 和 s, 有

$$A^r C^s \geqslant B^r D^s.$$

推论 7.3.1 若 A 和 B 为同阶半正定 Hermite 阵, $A \geqslant B \geqslant 0$. 则 $A^r \geqslant B^r$, 对一切 $r \in [0, 1]$ 成立. 特别 $A^{1/2} \geqslant B^{1/2}$.

§7.4 主 子 阵

设 $A = (a_{ij})$ 为 $n \times n$ 方阵, 我们用 $A\begin{pmatrix} i_1 & \cdots & i_p \\ \vdots & & \vdots \\ i_1 & \cdots & i_p \end{pmatrix}$ 表示由 A 的第 $i_1 \cdots i_p$ 行和第 $i_1 \cdots i_p$ 列交叉处元素组成的子方阵, 即主子阵. 下面的定理表明：正定阵的逆矩阵的任一主子阵在 Löwner 意义下总是不低于原方阵对应的主子阵的逆.

定理 7.4.1 设 A 为 $n \times n$ 正定阵 Hermite 阵, 则

$$(A^{-1})\begin{pmatrix} i_1 & \cdots & i_p \\ \vdots & & \vdots \\ i_1 & \cdots & i_p \end{pmatrix} \geqslant \left(A\begin{pmatrix} i_1 & \cdots & i_p \\ \vdots & & \vdots \\ i_1 & \cdots & i_p \end{pmatrix}\right)^{-1}.$$

证明 依定理 1.4.2(4), 将 A 的行进行互换, 同时将对应的列做同样的互换, 所得的方阵仍为正定阵, 故我们不妨假设 $i_1 = 1, \cdots, i_p = p$, 且把 A 分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

其中 A_{11} 为 $p \times p$ 方阵. 记

$$A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}.$$

利用分块矩阵求逆公式 (见定理 1.7.11), 有

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

因

$$A > 0 \Longrightarrow A_{22} > 0 \Longrightarrow A_{12}A_{22}^{-1}A_{21} \geqslant 0,$$

于是

$$A_{11} \geqslant A_{11} - A_{12}A_{22}^{-1}A_{21}.$$

因 $A_{11} > 0$, 由推论 7.2.2 得 $A^{11} \geqslant A_{11}^{-1}$, 即

$$A^{-1}\begin{pmatrix} 1 & \cdots & p \\ \vdots & & \vdots \\ 1 & \cdots & p \end{pmatrix} \geqslant \left(A\begin{pmatrix} 1 & \cdots & p \\ \vdots & & \vdots \\ 1 & \cdots & p \end{pmatrix}\right)^{-1}.$$

定理证毕.

下节我们将给出这个定理的另一种证明.

在定理中, 令 $p = 1$, 我们得到

推论 7.4.1 设 $A = (a_{ij})$ 为 $n \times n$ 正定阵, $A^{-1} = (a^{ij})$. 则

$$a^{ii} \geqslant \frac{1}{a_{ii}}, \qquad i = 1, \cdots, n.$$

§7.5 Cauchy-Schwarz 不等式的矩阵形式

在 §1.9, 我们已经证明, 对 $n \times n$ 正定 Hermite 阵 A 及任意两个 $n \times 1$ 向量 x 和 y, 有

$$|x^*y|^2 \leqslant x^*Ax \cdot y^*A^{-1}y,$$

特别, 当 $x=y$, $x^*x=1$ 时, 上式可以改写为

$$x^*A^{-1}x \geqslant (x^*Ax)^{-1}.$$

下面的定理把这个不等式推广到 x 为矩阵的情形, 它是由 Marshall 和 Olkin (1990) 证明的.

定理 7.5.1 设 A 为 $n\times n$ 正定 Hermite 阵, X 为 $n\times k$ 矩阵, 满足 $X^*X=I_k$. 则

$$X^*A^{-1}X \geqslant (X^*AX)^{-1}.$$

证明 设 Q 为 $n\times k$ 矩阵, 且 $r(Q)=k$, 则 $I_n - Q(Q^*Q)^{-1}Q^* \geqslant 0$. 设 P 为任意 $n\times l$ 矩阵, 于是

$$P^*(I_n - Q(Q^*Q)^{-1}Q^*)P \geqslant 0,$$

等价地

$$P^*P \geqslant P^*Q(Q^*Q)^{-1}Q^*P.$$

取 $P=A^{-1/2}X$, $Q=A^{1/2}X$, 上式则变为

$$X^*A^{-1}X \geqslant (X^*AX)^{-1}.$$

证毕.

注 1 定理 7.4.1 可以很容易从本定理推出. 事实上取 $X=(e_{i_1},\cdots,e_{i_p})$, 这里 $1\leqslant i_1\leqslant\cdots\leqslant i_p\leqslant n$, e_i 表示第 i 个元素为 1 其余元素皆为零的 $n\times 1$ 向量, 便得到定理 7.5.1.

注 2 Baksalary 和 Puntanen(1991) 把定理 7.5.1 推广到 A 是半正定 Hermite 阵的情形. 他们的结果是:

设 A 为 $n\times n$ 半正定 Hermite 阵, X 为 $n\times t$ 矩阵, 使得 $X^*A(A^*A)^-A^*X$ 为幂等阵, 则

$$X^*A^+X \geqslant (X^*AX)^+.$$

§7.6 Kantorovich 不等式的矩阵形式

在 §4.10, 我们证明了如下的 Kantorovich 不等式

$$\frac{x^*Axx^*A^{-1}x}{(x^*x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n},$$

其中 A 为 $n\times n$ 正定 Hermite 阵, λ_1 和 λ_n 分别为 A 的最大和最小特征值. 若 $x^*x=1$, 则上面的不等式可以改写为

$$x^*A^{-1}x \leqslant \frac{(\lambda+\lambda_n)^2}{4\lambda_1\lambda_n}(x^*Ax)^{-1}.$$

Marshall 和 Olkin(1990) 把这个不等式推广到 x 为矩阵的情形. 为了证明他们的结果. 我们需要如下引理.

引理 7.6.1　设 $a>0$, 则对任意的 $x\in[a,b]$, 总有

$$\frac{1}{x}\leqslant\frac{a+b}{ab}-\frac{x}{ab}.\tag{7.6.1}$$

证明　因为函数 $f(x)=x^{-1}$ 是 $[a,b]$ 上的凸函数, 于是对任意的 $\alpha\in[0,1]$, 有

$$f(\alpha a+(1-\alpha)b)\leqslant\alpha f(a)+(1-\alpha)f(b).$$

即

$$\frac{1}{\alpha a+(1-\alpha)b}\leqslant\frac{\alpha}{a}+\frac{1-\alpha}{b}.\tag{7.6.2}$$

注意到, 对任意 $x\in[a,b]$, 总存在 $\alpha\in[0,1]$, 将 x 表为 $x=\alpha a+(1-\alpha)b$, 由此解得

$$\alpha=\frac{x-b}{a-b},$$

代入 (7.6.2) 的右边, 整理便得到 (7.6.1) 的右边. 证毕.

定理 7.6.1　设 A 为 $n\times n$ 正定 Hermite 阵, X 为 $n\times t$ 矩阵, 满足 $X^*X=I_t$. 则

$$X^*A^{-1}X\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}.\tag{7.6.3}$$

其中 λ_1 和 λ_n 分别为 A 的最大和最小特征值.

证明　将 A 分解为 $A=U\Lambda U^*$, 这里 U 为 $n\times n$ 酉阵, $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_n)$, $\lambda_1\geqslant\cdots\geqslant\lambda_n>0$. 应用引理 7.6.1 得

$$\frac{1}{\lambda_i}\leqslant\frac{\lambda_1+\lambda_n}{\lambda_1\lambda_n}-\frac{\lambda_i}{\lambda_1\lambda_n},\qquad i=1,\cdots,n,$$

于是

$$\Lambda^{-1}\leqslant\frac{\lambda_1+\lambda_n}{\lambda_1\lambda_n}I_n-\frac{1}{\lambda_1\lambda_n}\Lambda.$$

用 X^*U 和 U^*X 分别左乘和右乘上式两边, 我们得到

$$X^*A^{-1}X\leqslant\frac{\lambda_1+\lambda_n}{\lambda_1\lambda_n}I_n-\frac{1}{\lambda_1\lambda_n}X^*AX,$$

上式右边可以改写为

$$\begin{aligned}&\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}-\frac{1}{\lambda_1\lambda_n}\left[\frac{(\lambda_1+\lambda_n)^2}{4}(X^*AX)^{-1}\right.\\&\left.-(\lambda_1+\lambda_n)I_n+X^*AX\right]\\=&\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}-\frac{1}{\lambda_1\lambda_n}\left(\frac{\lambda_1+\lambda_n}{2}(X^*AX)^{-1/2}-(X^*AX)^{1/2}\right)^2\\\leqslant&\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}.\end{aligned}$$

证毕.

在上面定理中, X 的列向量是任意 t 个标准正交化向量. 如果我们对它们加上一些约束条件, 那么不等式 (7.6.3) 还可以改进, 也就是说, 我们能够用更小的因子代替 $(\lambda_1+\lambda_n)^2/(4\lambda_1\lambda_n)$. 这就是下面的定理 7.6.2, 它是由本书作者之一和邵军证明的 (见 Wang Songgui 和 Shao Jun(1992)).

定理 7.6.2 设 $\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 $n\times n$ 实对称正定阵 A 的特征值, $\varphi_1,\cdots,\varphi_n$ 为对应的标准正交化特征向量, X 为 $n\times t$ 矩阵, 满足 $X'X=I_t$. 若存在 $1\leqslant i_1<\cdots<i_k\leqslant n$, 使得 $\mathscr{M}(X)\subset\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$. 则

$$X'A^{-1}X\leqslant\frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}(X'AX)^{-1}. \tag{7.6.4}$$

(7.6.4) 将作为一个更一般定理的推论而导出. 因为这个更一般定理的证明需要使用线性模型参数估计相对效率的概念, 所以我们留在第九章来讨论 (见 §9.3). 在那里, 我们还将给出 (7.6.4) 的若干统计应用.

注 对任意 $1\leqslant i_1<\cdots<i_k\leqslant n$, 容易验证

$$\frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}. \tag{7.6.5}$$

所以, (7.6.4) 是对 (7.6.3) 的一个改进. 但需说明, 这个改进仅在条件 $\mathscr{M}(X)\subset\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$ 成立时才成立.

§7.7 Wielandt 不等式的矩阵形式

在 §4.9, 我们讨论了如下的 Wielandt 不等式: 设 A 为 n 阶正定 Hermite 阵, λ_1 和 λ_n 为 A 的最大和最小特征值. 则对任意一对正交向量 x 和 y, 总有

$$|x^*Ay|^2\leqslant\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2x^*Ax\cdot y^*Ay. \tag{7.7.1}$$

不难看出, 这个不等式是 Cauchy-Schwarz 不等式的一种改进. 本节我们把它们推广到矩阵形式, 即 x 和 y 都为矩阵的情形.

我们先证明一个引理.

引理 7.7.1 设 A 为 n 阶正定 Hermite 阵, X 为 $n\times p$ 矩阵, $p\leqslant n$. 则

$$X(X^*A^{-1}X)^-X^*=A-AN(NAN)^-NA,$$

这里 $N=I-P_X$, $P_X=X(X^*X)^-X^*$.

证明 记 $D=X(X^*A^{-1}X)^-X^*-A+AN(NAN)^-NA$. 我们只需证明 $D=0$. 因为

$$D^*A^{-1}D=X(X^*A^{-1}X)^-X^*A^{-1}X(X^*A^{-1}X)^-X^*-2X(X^*A^{-1}X)^-X^*$$

$$-2AN(NAN)^-NA+AN(NAN)^-NAN(NAN)^-NA+A,$$

注意到上式第一项和第四项都与所包含的广义逆的选择无关, 因此我们将此处的广义逆用 Moore-Penrose 广义逆取代, 就可以证明它们分别等于 $X(X^*A^{-1}X)^-X^*$ 和 $AN(NAN)^-NA$, 据此证得

$$D^*A^{-1}D=-D.$$

再用 $A^{-1/2}$ 左乘和右乘上式, 并记 $M=A^{-1/2}DA^{-1/2}$, 便有 $M^2=-M$, 这表明 $M=0$, 因而必有 $D=0$. 引理得证.

下面的定理就是 Wielandt 不等式的矩阵形式, 它是由本书作者之一和叶伟彰证明的 (Wang, Ip1999).

定理 7.7.1 设 A 为 n 阶正定 Hermite 阵, λ_1 和 λ_n 为 A 的最大和最小特征值, X 和 Y 分别为 $n\times p$ 和 $n\times q$ 矩阵, 且满足 $X^*Y=0$. 则

$$X^*AY(Y^*A^{-1}Y)^-Y^*AX\leqslant\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2X^*AX. \tag{7.7.2}$$

证明 用 $X(X^*X)^-$ 和 $(X^*X)^-X^*$ 分别左乘和右乘 (7.7.2) 得

$$P_XAY(Y^*A^{-1}Y)^-Y^*AP_X\leqslant\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2P_XAP_X, \tag{7.7.3}$$

这里 P_X 的定义同引理. 很显然, (7.7.2) 和 (7.7.3) 是等价的. 于是我们将证明后者. 但是, 这意味着要证明, 对任意 $n\times 1$ 向量 c, 有

$$c^*P_XAY(Y^*A^{-1}Y)^-Y^*AP_Xc\leqslant\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2c^*P_XAP_Xc,$$

上式可改写为

$$\frac{c^*P_XAY(Y^*A^{-1}Y)^-Y^*AP_Xc}{c^*P_XAP_Xc} \leqslant \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2. \tag{7.7.4}$$

注意到 λ_1 和 λ_n 皆为正数, 因而

$$\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n} \leqslant 1,$$

于是 (7.7.4) 可进一步改写为

$$\frac{c^*P_XAP_Xc-c^*P_XAY(Y^*A^{-1}Y)^-Y^*AP_Xc}{c^*P_XAP_Xc} \geqslant 1-\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 = \left(\frac{4\lambda_1\lambda_n}{\lambda_1+\lambda_n}\right)^2. \tag{7.7.5}$$

再利用不等式

$$N(NAN)^-N \geqslant Y(Y^*A^{-1}Y)^-Y,$$

这里 N 的定义同引理 7.7.1. 因而 (7.7.5) 成立的一个充分条件是

$$\frac{c^*P_X(A-AN(NAN)^-NA)P_Xc}{c^*P_XAP_Xc} \geqslant \left(\frac{4\lambda_1\lambda_n}{\lambda_1+\lambda_n}\right)^2.$$

又由引理 7.7.1, 上式可变为

$$\frac{c^*X(X^*A^{-1}X)^-X^*c}{c^*P_XAP_Xc} \geqslant \left(\frac{4\lambda_1\lambda_n}{\lambda_1+\lambda_n}\right)^2. \tag{7.7.6}$$

为了证明上式, 我们对 X 作奇异值分解

$$X = H\Delta G^*,$$

其中 H 为 $n\times r$ 矩阵, G 为 $p\times r$ 矩阵, 满足 $H^*H=I_r$ 和 $G^*G=I_r$, 且 Δ 为 $r\times r$ 对角阵, 其对角元皆为正数, r 为 A 的秩 (关于奇异值分解, 参阅定理 1.5.4). 应用定理 1.7.9, 容易证明

$$X(X^*A^{-1}X)^-X^* = H(H^*A^{-1}H)^-H^*.$$

注意到 $P_X = HH^*$. 记 $t = H^*c$, 则 (7.7.6) 变形为

$$\frac{t^*(H^*A^{-1}H)^{-1}t}{t^*H^*AHt} \geqslant \left(\frac{4\lambda_1\lambda_n}{\lambda_1+\lambda_n}\right)^2. \tag{7.7.7}$$

由 Cauchy-Schwarz 不等式可知

$$(t^*t)^2 \leqslant t^*H^*A^{-1}Ht\cdot t^*(H^*A^{-1}H)^{-1}t,$$

于是

$$\frac{t^*(H^*A^{-1}H)^{-1}t}{t^*H^*AHt} \geqslant \frac{(t^*t)^2}{t^*H^*AHt \cdot t^*H^*A^{-1}Ht} = \frac{(u^*u)^2}{u*A^{-1}u \cdot u*A^{-1}u}. \tag{7.7.8}$$

这里 $u = Ht$, 满足 $u^*u = t^*t$. 应用 Kantorovich 不等式 (见定理 4.10.1), 我们有

$$\frac{(u^*u)^2}{u*A^{-1}u \cdot u*A^{-1}u} \geqslant \left(\frac{4\lambda_1\lambda_n}{\lambda_1+\lambda_n}\right)^2. \tag{7.7.9}$$

综合 (7.7.6) ~ (7.7.9), 便证明了不等式 (7.7.2). 定理证毕.

推论 7.7.1　设 A 为 n 阶正定 Hermite 阵, 将其分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

则

$$A_{12}A_{22}^{-1}A_{21} \leqslant \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 A_{11}, \tag{7.7.10}$$

这里 λ_1 和 λ_n 为 A 的最大和最小特征值.

证明　设 A_{11} 和 A_{22} 分别为 p 阶和 q 矩阵, $p+q=n$. 则在定理中取 X 和 Y 分别为

$$X = \begin{pmatrix} I_p \\ 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 \\ I_q \end{pmatrix},$$

便得到 (7.7.10). 推论证毕.

§7.8　凸函数的矩阵不等式

本节我们先介绍凸函数的矩阵不等式. 据此我们可以获得许多重要的不等式, 如 Kantorovich 不等式, Jensen 不等式的矩阵形式以及 Hadamard 不等式.

由定理 1.4.1 知, 如果 A 为 n 阶 Hermite 阵, 则总存在一个酉阵 U, 使得

$$A = U^*\mathrm{diag}(\lambda_1, \cdots, \lambda_n)U, \tag{7.8.1}$$

这里 $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值. 这时定义 $f(A)$ 为

$$f(A) = U^*\mathrm{diag}(f(\lambda_1), \cdots, f(\lambda_n))U.$$

如果 $F(t) = F(f(t), g(t))$, 则 $F(A) = F(f(A), g(A))$, $F(A, B)$ 表示两矩阵变量的矩阵函数.

定理 7.8.1 设 A 为 n 阶 Hermite 阵, X 为 $n\times l$ 矩阵, 满足 $X^*X=I_l$. 若 $f(t)$ 为 $[\lambda_1,\lambda_n]$ 上的一个连续凸函数, 则

$$X^*f(A)X\leqslant\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}I_l+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}X^*AX, \tag{7.8.2}$$

其中 λ_1 和 λ_n 分别为 A 的最大和最小特征值.

证明 因为函数 $f(t)$ 为 $[\lambda_1,\lambda_n]$ 上的连续凸函数, 于是对任意的 $a\in[\lambda_1,\lambda_n]$, 有

$$f(a)\leqslant\frac{\lambda_1-a}{\lambda_1-\lambda_n}f(\lambda_n)+\frac{a-\lambda_n}{\lambda_1-\lambda_n}f(\lambda_1).$$

结合 (7.8.1), 于是

$$f(A)\leqslant\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}I_n+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}A.$$

由 $X^*X=I_l$, 定理证毕.

定理 7.8.2 在定理 7.8.1 的假设下, 设 D 为 $f(t)$ 在 $[\lambda_1,\lambda_n]$ 上所对应的值域子区间, 二元实函数 $F(u,v)$ 在 $D\times D$ 上连续, 且关于变量 u 是单调增函数, 则

$$F\Big[X^*f(A)X,f(X^*AX)\Big]\leqslant\Big\{\max_{\theta\in[0,1]}F\Big[\theta f(\lambda_n)+(1-\theta)f(\lambda_n),f(\theta\lambda_n+(1-\theta)\lambda_n)\Big]\Big\}I_l. \tag{7.8.3}$$

证明 由 (7.8.2) 和 $F(u,v)$ 关于变量 u 单调性, 我们立得

$$F\Big[X^*f(A)X,f(X^*AX)\Big]\leqslant F\left[\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}I_l+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}X^*AX,f(X^*AX)\right].$$

记

$$K=\max_{\lambda_n\leqslant t\leqslant\lambda_1}F\left[\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}t,\ f(t)\right],$$

则

$$F\left[\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}t,\ f(t)\right]\leqslant K,$$

由于 $\lambda_n I\leqslant X^*AX\leqslant\lambda_1 I$, 类似于 Mond 和 Pečarić (1995,b) 的证明, 有

$$F\left[\frac{\lambda_1 f(\lambda_n)-\lambda_n f(\lambda_1)}{\lambda_1-\lambda_n}+\frac{f(\lambda_1)-f(\lambda_n)}{\lambda_1-\lambda_n}X^*AX,\ f(X^*AX)\right]\leqslant KI. \tag{7.8.4}$$

令 $\theta=(\lambda_1-t)/(\lambda_1-\lambda_n)$, 则 $t=\theta\lambda_n+(1-\theta)\lambda_n$, $0\leqslant\theta\leqslant 1$, 因此

$$K=\Big\{\max_{\theta\in[0,1]}F\Big[\theta f(\lambda_n)+(1-\theta)f(\lambda_n),f(\theta\lambda_n+(1-\theta)\lambda_n)\Big]\Big\}. \tag{7.8.5}$$

由 (7.8.4) 和 (7.8.5) 便得 (7.8.6). 定理证毕.

将定理 7.8.2 中的 $F(u,v)$ 用 $-F(u,v)$ 取代, 用同样的方法我们可证得下列定理.

定理 7.8.3 若定理 7.8.2 中的二元实函数 $F(u,v)$ 关于变量 u 是单调减函数, 则

$$F\Big[X^*f(A)X, f(X^*AX)\Big] \leqslant \Big\{\min_{\theta\in[0,1]} F\Big[\theta f(\lambda_n)+(1-\theta)f(\lambda_n), f(\theta\lambda_n+(1-\theta)\lambda_n)\Big]\Big\} I_l. \tag{7.8.6}$$

显然, $F(u,v)=u-v$ 和 $F(u,v)=u/v$ 皆为关于变量 u 是单调增函数, 应用这两个函数, 类似于 Mond 和 Pečarić(1995,b) 的证明, 我们可得到如下结果.

定理 7.8.4 在定理 7.8.1 的假设下, 若 $f(t)$ 是二阶可导的严格凸函数, 且在 $[\lambda_1,\lambda_n]$ 上取值恒为正数 (或恒为负数), 则存在某个实数 $\gamma>1$(或 $0<\gamma<1$), 使得

$$X^*f(A)X \leqslant \gamma f(X^*AX), \tag{7.8.7}$$

这里 γ 仅依赖于 f,λ_1 和 λ_n. 令 $\mu=(f(\lambda_1)-f(\lambda_n))/(\lambda_1-\lambda_n)$. 如果 $\mu=0$, 设 t_0 为方程 $f'(t)=0$ 在 (λ_1,λ_n) 上的唯一解, 则 $\gamma=f(\lambda_n)/f(t_0)$ 满足不等式 (7.8.7). 如果 $\mu\neq 0$, 设 t_0 为方程

$$\mu f(t)-f'(t)(f(\lambda_n)+\mu(t-\lambda_n))=0$$

在 (λ_1,λ_n) 上的唯一解, 则 $\gamma=\mu/f'(t_0)$ 满足不等式 (7.8.7).

定理 7.8.5 在定理 7.8.1 的假设下, 若 $f(t)$ 可导, $f'(t)$ 在 $[\lambda_1,\lambda_n]$ 上是严格增函数, 则

$$X^*f(A)X \leqslant \gamma I + f(X^*AX), \tag{7.8.8}$$

这里 γ 仅依赖于 f,λ_1 和 λ_n. 设 t_0 为方程 $f'(t)=\mu$ 在 (λ_1,λ_n) 上的唯一解, 则 $\gamma=f(\lambda_n)-f(t_0)+\mu(\lambda_1-\lambda_n)$ 满足不等式, 这里 $\mu=(f(\lambda_1)-f(\lambda_n))/(\lambda_1-\lambda_n)$.

令 $f(t)=t^p(t>0)$. 显然当 $p>1$, 或$p<0$ 时, $f(t)$ 为凸函数, 当 $0<p<1$ 时, $-f(t)$ 为凸函数. 因此依据定理 7.8.4 和 7.8.5, 便得到以下两个重要的不等式.

推论 7.8.1 设 A 为 n 阶正定 Hermite 阵, X 为 $n\times l$ 矩阵, 满足 $X^*X=I_l$, 则对任意的实数 $p>1$或$p<0$,

$$X^*A^pX \leqslant \gamma(X^*AX)^p, \tag{7.8.9}$$

这里

$$\gamma=\frac{(p-1)^{p-1}}{p^p}\frac{(\lambda_1^p-\lambda_n^p)^p}{(\lambda_1-\lambda_n)(\lambda_n\lambda_1^p-\lambda_1\lambda_n^p)^{p-1}}. \tag{7.8.10}$$

推论 7.8.2 设 A 为 n 阶正定 Hermite 阵, X 为 $n\times l$ 矩阵, 满足 $X^*X=I_l$, 则对任意的实数 p ($p>1$ 或 $p<0$)

$$X^*A^pX-(X^*AX)^p \leqslant \gamma I, \tag{7.8.11}$$

这里

$$\gamma=\lambda_n^p-\Big(\frac{1}{p}\frac{\lambda_1^p-\lambda_n^p}{\lambda_1-\lambda_n}\Big)^{p/(p-1)}+\frac{\lambda_1^p-\lambda_n^p}{\lambda_1-\lambda_n}\left[\Big(\frac{1}{p}\frac{\lambda_1^p-\lambda_n^p}{\lambda_1-\lambda_n}\Big)^{1/(p-1)}-\lambda_n\right], \tag{7.8.12}$$

且对任意的实数 $0<p<1$, 不等式 (7.8.11) 的不等号变向.

注 1 在推论 7.8.1 的假设下, 若 $0<p<1$, 则令 $q=1/p>1$, $B=A^p$. 于是 $(X^*A^pX)^{1/p}=(X^*BX)^q$, $X^*AX=X^*B^qX$. 应用 (7.8.9) 容易验证

$$X^*AX\leqslant\tau^{1/p}(X^*A^pX)^{1/p},$$

即

$$(X^*AX)^p\leqslant\tau X^*A^pX, \tag{7.8.13}$$

其中

$$\tau=p^p(1-p)^{1-p}\frac{(\lambda_1-\lambda_n)}{(\lambda_1^p-\lambda_n^p)^p(\lambda_n^p\lambda_1-\lambda_1^p\lambda_n)^{(1-p)}}.$$

应用 (7.8.9) 和 (7.8.11), 我们可得到许多重要的矩阵不等式. 例如, 令 (7.8.9) 中的 $p=-1$, 便得到 Kantorovich 不等式

$$X^*A^{-1}X\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}. \tag{7.8.14}$$

而由 (7.8.11) 得

$$X^*A^{-1}X-(X^*AX)^{-1}\leqslant\frac{(\sqrt{\lambda_1}+\sqrt{\lambda_n})^2}{\lambda_1\lambda_n}I_n. \tag{7.8.15}$$

令 (7.8.9) 和 (7.8.11) 中的 $p=2$, 便得到 Liu 和 Neudecker(1996) 的两个 Kantorovich 型不等式

$$X^*A^2X\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^2,$$

$$X^*A^2X-(X^*AX)^2\leqslant\frac{1}{4}(\lambda_1-\lambda_n)^2I.$$

令 (7.8.13) 和 (7.8.12) 中 $p=1/2$, 得

$$(X^*AX)^{1/2}\leqslant\frac{\sqrt{\lambda_1}+\sqrt{\lambda_n}}{2\sqrt{\lambda_1\lambda_n}}(X^*A^{1/2}X),$$

$$(X^*AX)^{1/2}-X^*A^{1/2}X\leqslant\frac{(\sqrt{\lambda_1}-\sqrt{\lambda_n})^2}{4(\sqrt{\lambda_1}+\sqrt{\lambda_n})}I.$$

令 $A = \mathrm{diag}(A_1, \cdots, A_k)$, $X^* = (X_1^*, \cdots, X_k^*)$, 其中每个 A_i 和 X_i 分别为 n 阶 Hermite 阵和 $n \times l$ 的矩阵, 于是有

$$X^*AX = \sum_{j=1}^{k} X_j^* A_j X_j.$$

若 $\sum_{j=1}^{k} X_j^* X_j = I_l$, 每个 A_i 为 n 阶正定阵, 且其特征值都在 $[m, M]$ 上, 则由 (7.8.9) 和 (7.8.11), 我们立得, 对任意整数 $p \neq 0, 1$, 有

$$\sum_{j=1}^{k} X_j^* A_j X_j \leqslant \gamma_1 \left(\sum_{j=1}^{k} X_j^* A_j X_j \right)^p, \tag{7.8.16}$$

$$\sum_{j=1}^{k} X_j^* A_j X_j - \left(\sum_{j=1}^{k} X_j^* A_j X_j \right)^p \leqslant \gamma_2 I, \tag{7.8.17}$$

这里, γ_1 和 γ_2 分别为 (7.8.10) 和 (7.8.12), 只需将其中的 λ_1 和 λ_n 分别替换为 M 和 m. 这是 Jensen 不等式的两个重要矩阵形式.

注 2　只要 $\sum_{j=1}^{k} X_j^* X_j = I_l$, 且 A_i 满足本节相应定理和推论的条件, 则用 $\sum_{j=1}^{k} X_j^* A_j X_j$, M 和 m 分别取代本节所有定理和推论中的 X^*AX, λ_1 和 λ_n, 便得到相应的 Jensen 不等式的矩阵形式, 见 Mond 和 Pečarić(1997).

注 3　令注 2 中的每个 X_i 皆为 $n \times 1$ 向量, 即 $X_i = x_i$, $A_j = A$. 只要 $\sum_{j=1}^{k} \|x_j\|^2 = 1$, 则对任意整数 $p \neq 0, 1$, 有

$$\frac{\sum_{j=1}^{k} x_j^* A^p x_j}{\left(\sum_{j=1}^{k} x_j^* A x_j\right)^p} \leqslant \frac{(p-1)^{p-1}}{p^p} \frac{(\lambda_1^p - \lambda_n^p)^p}{(\lambda_1 - \lambda_n)(\lambda_n \lambda_1^p - \lambda_1 \lambda_n^p)^{p-1}}.$$

这便是 Ky Fan 证明的另一个不等式, 见 Mond 和 Pečarić(1995, c).

以上结论还可以被推广到 A 为半正定阵的情形, 这里我们仅叙述主要结果, 其证明读者可参阅 Malamud(2001). 设 $r(A) = r < n$, 则存在 $r \times n$ 的矩阵 H, 使得

$$A = H^* \mathrm{diag}(\lambda_1, \cdots, \lambda_r) H,$$

其中 $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ 为 A 的非零的特征值, H 满足 $HH^* = I_r$, 于是 $P_A = H^*H$. 易见

$$A^+ = H^* \mathrm{diag}(1/\lambda_1, \cdots, 1/\lambda_r) H.$$

若 X^*P_AX 为幂等阵, 则

$$X^*A^+X \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} (X^*AX)^+,$$

$$X^*A^2X \leqslant \frac{(\lambda_1+\lambda_r)^2}{4\lambda_1\lambda_r}(X^*AX)^2,$$

$$0 \leqslant X^*AX - (X^*A^+X)^+ \leqslant \left(\sqrt{\lambda_1}-\sqrt{\lambda_r}\right)^2 X^*P_AX,$$

$$0 \leqslant X^*A^+X - (X^*AX)^+ \leqslant \frac{(\sqrt{\lambda_1}-\sqrt{\lambda_r})^2}{\lambda_1\lambda_r} X^*P_AX,$$

$$X^*A^2X - (X^*AX)^2 \leqslant (\sqrt{\lambda_1}-\sqrt{\lambda_r})^2 X^*P_AX,$$

$$(X^*A^2X)^{1/2} - X^*AX \leqslant \frac{(\lambda_1-\lambda_r)^2}{4(\lambda_1+\lambda_r)} X^*P_AX.$$

由 (7.8.9) 和 (7.8.11), 我们还可以得到一些关于 Hadamard 乘积的不等式, 这将在下一节讨论.

此外, 文献中还研究了两个正定 Hermite 阵凸组合的逆的不等式, 设 A 和 B 为 n 阶正定 Hermite 阵, $0 \leqslant \alpha \leqslant 1$, 则

$$\left[\alpha A + (1-\alpha)B\right]^{-1} \leqslant \left[\alpha A^{-1} + (1-\alpha)B^{-1}\right].$$

Mond 和 Pečarić(1995, b) 进一步给出了凸组合的逆的下界, 其结果归纳成见如下定理.

定理 7.8.6 设 A 和 B 为 n 阶正定 Hermite 阵, $0 \leqslant \alpha \leqslant 1$, 则

$$\left[\alpha A + (1-\alpha)B\right]^{-1} \geqslant \tau_1\left[\alpha A^{-1} + (1-\alpha)B^{-1}\right], \tag{7.8.18}$$

$$\left[\alpha A^{-1} + (1-\alpha)B^{-1}\right] - \left[\alpha A + (1-\alpha)B\right]^{-1} \leqslant \tau_2 A^{-1}, \tag{7.8.19}$$

这里

$$\tau_1 = 4\min\left\{\frac{\mu_1}{(1+\mu_1)^2}, \frac{\mu_n}{(1+\mu_n)^2}\right\}, \qquad \tau_2 = \min_i \frac{(\sqrt{\mu_i}-1)^2}{\mu_i},$$

其中 μ_1 和 μ_n 分别为 BA^{-1} 的最大、最小特征值.

证明 用 $A^{1/2}$ 左乘和右乘 (7.8.18) 和 (7.8.19) 得

$$\left[\alpha I_n + (1-\alpha)A^{-1/2}BA^{-1/2}\right]^{-1} \geqslant \tau_1\left[\alpha I_n + (1-\alpha)A^{1/2}B^{-1}A^{1/2}\right], \tag{7.8.20}$$

$$\left[\alpha I_n + (1-\alpha)A^{1/2}B^{-1}A^{1/2}\right] - \left[\alpha I_n + (1-\alpha)A^{-1/2}BA^{-1/2}\right]^{-1} \leqslant \tau_2 I_n, \tag{7.8.21}$$

令 $\mu_1 \geqslant \cdots \geqslant \mu_n$ 为 BA^{-1} 的特征值, 记 $D = \mathrm{diag}(\mu_1, \cdots, \mu_n)$, 则存在一酉矩阵 H, 使得

$$A^{-1/2}BA^{-1/2} = H^*DH,$$

因此, (7.8.20) 和 (7.8.21) 分别等价于

$$\frac{1}{\alpha+(1-\alpha)\mu_i} \geqslant \tau_1\Big(\alpha+(1-\alpha)\frac{1}{\mu_i}\Big), \quad i=1,2,\cdots,n,$$

$$\frac{1}{\alpha+(1-\alpha)\mu_i} - \Big(\alpha+(1-\alpha)\frac{1}{\mu_i}\Big) \geqslant \tau_2, \quad i=1,2,\cdots,n.$$

令 $x=(\alpha,1-\alpha)'$, $D_i=\mathrm{diag}(1,\mu_i)$, 则

$$\alpha+(1-\alpha)\mu_i = x'D_ix, \quad \alpha+(1-\alpha)\frac{1}{\mu_i} = x'D^{-1}x,$$

利用 Kantorovich 不等式 (4.10.4) 或 (7.8.14), 立得

$$\frac{1}{\alpha+(1-\alpha)\mu_i} \geqslant \frac{4\mu_i}{(1+\mu_i)^2}\Big(\alpha+(1-\alpha)\frac{1}{\mu_i}\Big), \quad i=1,2,\cdots,n, \tag{7.8.22}$$

由 $f(t)=t/(1+t)^2$ 的单调性, 我们易证

$$\frac{4\mu_i}{(1+\mu_i)^2} \geqslant \tau_1 = 4\min\left\{\frac{\mu_1}{(1+\mu_1)^2}, \frac{\mu_n}{(1+\mu_n)^2}\right\},$$

故 (7.8.18) 得证. 利用 (7.8.15), 我们有

$$\frac{1}{\alpha+(1-\alpha)\mu_i} - \Big(\alpha+(1-\alpha)\frac{1}{\mu_i}\Big) \geqslant \frac{(\sqrt{\mu_i}-1)^2}{\mu_i} \geqslant \tau_2, \quad i=1,2,\cdots,n.$$

因此 (7.8.19) 得证. 定理证毕.

§7.9 Hadamard 乘积

在 §1.10, 我们引进了矩阵的 Hadamard 乘积, 并讨论了一些初等性质. 本节将证明 Hadamard 乘积有关偏序的几个结果.

为了查阅的方便, 我们先把定理 1.10.1 转述在这里.

定理 7.9.1 设 A 和 B 皆为 Hermite 阵.

(1) 设 $A\geqslant 0$, $B\geqslant 0$, 则 $A\circ B\geqslant 0$;

(2) 设 $A>0$, $B>0$, 则 $A\circ B>0$.

定理 7.9.2 设 A, B, C 和 D 皆为 Hermite 阵, $A\geqslant B$.

(1) 设 $C\geqslant 0$, 则 $A\circ C\geqslant B\circ C$;

(2) 设 $C\geqslant D\geqslant 0$, 则 $A\circ C\geqslant B\circ D$.

证明 (1) 因 $A\geqslant B$, 故 $A-B\geqslant 0$, 应用上一定理之 (1), 我们有 $(A-B)\circ C\geqslant 0$, 即 $(A\circ C)-(B\circ C)\geqslant 0$. 得证.

(2) 是 (1) 的直接推论.

定理 7.9.3 设 A 和 B 为两个 n 阶正定 Hermite 阵, 则

(1) $A^{-1} \circ B^{-1} \geqslant (A \circ B)^{-1}$;

(2) $A^{-1} \circ A^{-1} \geqslant (A \circ A)^{-1}$;

(3) $A^{-1} \circ A \geqslant I \geqslant (A^{-1} \circ A)^{-1}$.

证明 (1) 根据 Kronecker 乘积的性质 (7), 由 A 和 B 可逆知, $A \otimes B$ 也可逆, 且 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

记 $J = \{1, n+2, 2n+3, 3n+4, \cdots, n^2\}$, $C(J)$ 表示方阵 C 中对应于 J 的那些行和列组成子方阵. 利用定理 1.10.3 和定理 7.4.1, 我们有

$$\begin{aligned} A^{-1} \circ B^{-1} &= (A^{-1} \otimes B^{-1})(J) = (A \otimes B)^{-1}(J) \\ &\geqslant ((A \otimes B)(J))^{-1} \\ &= (A \circ B)^{-1}. \end{aligned}$$

(2) 是 (1) 的明显特例.

(3) 在 (1) 中, 取 $B = A^{-1}$, 得

$$A^{-1} \circ A \geqslant (A \circ A^{-1})^{-1} = (A^{-1} \circ A)^{-1}.$$

因为对一个可逆 Hermite 阵 M, $M \geqslant M^{-1} \Longleftrightarrow \lambda_i(M) \geqslant 1, i = 1, \cdots, n$. 于是, $M \geqslant I \geqslant M^{-1}$. 把此结论应用于 $A^{-1} \circ A$, (3) 得证.

应用归纳法, 可以证明

推论 7.9.1 设 A 为正定 Hermite 阵, 则

$$A^{-1} \circ A^{-1} \circ \cdots \circ A^{-1} \geqslant (A \circ A \circ \cdots \circ A)^{-1}.$$

对于矩阵的普通乘积, 我们有 $A^{-1}A = I$. 于是从定理 7.9.3 之 (3), 我们立即得到如下推论, 它给出了矩阵的 Hadamard 乘积与普通乘积的关系.

推论 7.9.2 设 A 为正定 Hermite 阵, 则

$$A^{-1} \circ A \geqslant A^{-1}A = I.$$

定理 7.9.3 给出了两同阶正定阵 Hadamard 乘积 $A^{-1} \circ B^{-1}$ 的下界, 下面我们将给出它的一个上界.

令 $X = (x_{ij})$ 为 $n^2 \times n$ 矩阵, 其中 $x_{i,j} = 1$, 如果 $i = (j-1)n + j$, 否则 $x_{i,j} = 0$. 显然有

$$X^*X = I_n, \quad A \circ B = X^*(A \otimes B)X.$$

注意到

$$A^p \circ B^p = X^*(A^p \otimes B^p)X = X^*(A \otimes B)^p X,$$

由推论 7.8.1 和 7.8.2, 我们可得到如下定理.

定理 7.9.4　设 A 和 B 为两个 n 阶的正定 Hermite 阵, 则对任意的非零整数 p, 有

$$A^p \circ B^p \leqslant \gamma_1 (A \circ B)^p, \tag{7.9.1}$$

$$A^p \circ B^p - (A \circ B)^p \leqslant \gamma_2 I_n, \tag{7.9.2}$$

这里 γ_1, γ_2 分别如 (7.8.10),(7.8.12) 所定义的 γ, 只需将其中对应的特征值替换为 $A \otimes B$ 的最大、最小特征值 λ_1 和 λ_{n^2}.

注意到 $A \otimes B$ 的特征值是 A 的特征值与 B 的特征值相乘所得的 n^2 个积, 因此若设 A, B 的最大、最小特征值分别 $\alpha_1, \alpha_n, \beta_1, \beta_n$, 则

$$\lambda_1 = \alpha_1 \beta_1, \quad \lambda_{n^2} = \alpha_n \beta_n.$$

分别令 $p = 2$, $p = -1$, 则有

$$A^2 \circ B^2 \leqslant \frac{(\alpha_1\beta_1 + \alpha_n\beta_n)^2}{4\alpha_1\beta_1\alpha_n\beta_n}(A \circ B)^2,$$

$$A^{-1} \circ B^{-1} \leqslant \frac{(\alpha_1\beta_1 + \alpha_n\beta_n)^2}{4\alpha_1\beta_1\alpha_n\beta_n}(A \circ B)^{-1},$$

$$A^2 \circ B^2 - (A \circ B)^2 \leqslant \frac{(\alpha_1\beta_1 + \alpha_n\beta_n)^2}{4} I_n,$$

$$A^{-1} \circ B^{-1} - (A \circ B)^{-1} \leqslant \frac{(\sqrt{\alpha_1\beta_1} - \sqrt{\alpha_n\beta_n})^2}{\alpha_1\beta_1\alpha_n\beta_n} I_n.$$

第8章　受　　控

受控 (majorization) 是向量之间的一种较弱的次序关系. 这个概念最早出现在 Hardy, Littlewood 和 Polya(1934) 的名著 "Inequalities" 一书中. 自那以后, 这方面的研究引起了许多学者的相当兴趣, 有关受控的理论及其应用获得了长足的发展, 受控与 Schur 函数的结合已成为导出不等式的一种有效方法. 在数学的诸多分支, 特别是在组合数学、几何学、数值分析和概率统计, 受控理论都有着许多的应用, 详细讨论见 Marshall 和 Olkin(1979).

本章我们仅限于讨论矩阵论中的受控关系. §8.1 引进了受控和表征它的一个重要工具 —— 双随机阵; §8.2 讨论了 Schur 函数的基本性质; 在接下来的 §8.3~§8.7, 研究了各种矩阵的对角元向量、特征值向量 (即特征值组成的向量) 以及奇异值向量之间的受控关系.

§8.1　基 本 概 念

众所周知, 对实轴上的任意两个数 x 和 y, 都存在着明确的大小关系, 或 $x \leqslant y$ 或 $x > y$, 二者必居其一. 但是, 对于 R^n 中的向量 x 和 y, 我们就不能这样简单地去比较了. 此时, 存在着多种比较 x 与 y 大小的方法. 最简单的是比较它们的长度 $||x||$ 和 $||y||$. 另一种是逐个分量进行比较, 若对 $x = (x_1, \cdots, x_n)'$ 和 $y = (y_1, \cdots, y_n)'$, $x_i = y, i = 1, \cdots, k, x_{k+1} < y_{k+1}$ 就称 $x \prec y$. 本节我们要讨论的受控也是描述 n 维向量大小的一种关系, 它不是从分量角度比较两个向量, 而是从它们的分量部分和来比较. 具体定义如下.

设 $x = (x_1, \cdots, x_n)' \in R^n$, $x_{(1)} \geqslant \cdots \geqslant x_{(n)}$ 表示向量 x 的各分量按递减次序排列.

定义 8.1.1　设 $x, y \in R^n$. 若

$$\sum_{i=1}^{k} x_{(i)} \leqslant \sum_{i=1}^{k} y_{(i)}, \qquad k = 1, \cdots, n-1, \tag{8.1.1}$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \tag{8.1.2}$$

则称 x 受控于 y, 或称 y 控制 x, 记为 $x \prec y$. 若

$$\sum_{i=1}^{k} x_{(i)} \leqslant \sum_{i=1}^{k} y_{(i)}, \qquad k=1,\cdots,n, \tag{8.1.3}$$

则称 x 弱受控于 y, 或称 y 弱控制 x, 记为 $x \prec_w y$. 容易验证, 受控关系具有下列性质：

(1) 自反性：$x \prec x$, $(x \prec_w x)$;

(2) 传递性：$x \prec y$, $y \prec z$, 则 $x \prec z$(若 $x \prec_w y$, $y \prec_w z$, 则 $x \prec_w z$);

(3) 置换不变性：设 $x \prec y$, 则对任意两个置换阵 π_1 和 π_2(所谓置换阵就是由单位阵经过行 (列) 互换而得到的矩阵. 因而一个置换阵就是每行每列有一个且只有一个元素为 1, 而其余元素为零的矩阵), 有 $\pi_1 x \prec \pi_2 y$(若 $x \prec_w y$, 有 $\pi_1 x \prec_w \pi_2 y$).

注 1　条件 (8.1.1) 等价于

$$\max_{1\leqslant j_1<\infty<j_k\leqslant n} \sum_{i=1}^{k} x_{j_i} \leqslant \max_{1\leqslant j_1<\cdots<j_k\leqslant n} \sum_{i=1}^{k} y_{j_i}, \quad k=1,\cdots,n-1.$$

注 2　在 §7.1, 我们引进了偏序 (partial ordering) 的概念. 请读者注意, 此处的受控并不是偏序, 因为它不满足偏序定义中的第三条, 即由 $x \prec y$, $y \prec x$, 我们不能推出 $x=y$. 例如, $x=(1,2,3)'$, $y=(3,2,1)'$, 就是一例. 从定义可以看出, 若 $x \prec y$ $y \prec x$ 则存在一个置换阵 π, 使得 $x=\pi y$. 为了区别起见, 在文献中, 把满足自反性和传递性的关系称为预序 (preordering). 因此, 受控和弱受控都是预序而非偏序.

注 3　若将定义 8.1.1 中的 R^n 用其子集 S 来代替, 则我们在一个子集 S 上定义了受控关系. 在后面的讨论中, 我们经常需要考虑在 $D=\{x=(x_1,\cdots,x_n)', x_1 \geqslant \cdots \geqslant x_n\}$ 上的受控. 附带说明一点, 若限制在 D 上, 受控和弱受控都是偏序.

下面是受控的几个例子.

例 8.1.1　$\left(\frac{1}{n},\cdots,\frac{1}{n}\right)' \prec \left(\frac{1}{n-1},\cdots,\frac{1}{n-1},0\right)' \prec \cdots \prec$

$$\prec \left(\frac{1}{2},\frac{1}{2},0,\cdots,0\right)' \prec (1,0,\cdots,0)'$$

实际上, 由于置换不变性, 上面所有零分量可以换到任何其他位置, 而保持受控关系不变.

例 8.1.2　$\left(\frac{1}{n},\cdots,\frac{1}{n}\right)' \prec (a_1,\cdots,a_n)' \prec (0,\cdots,0,1,0,\cdots,0)'$, 对一切满足 $a_i \geqslant 0$, $\sum_{i=1}^{n} a_i = 1$ 的 $(a_1,\cdots,a_n)'$ 成立.

为了进一步刻画受控关系, 我们需引进双随机阵的概念.

定义 8.1.2 设 $P=(p_{ij})$ 为 $n\times n$ 矩阵, 满足

$$p_{ij}\geqslant 0,\quad i,j=1,\cdots,n, \tag{8.1.4}$$

$$\sum_{j=1}^{n}p_{ij}=1,\quad i=1,\cdots,n, \tag{8.1.5}$$

$$\sum_{i=1}^{n}p_{ij}=1,\quad j=1,\cdots,n, \tag{8.1.6}$$

则称 P 为双随机阵. 若在 (8.1.5) 和 (8.1.6) 中, "$\leqslant$" 号成立, 则称 P 为次双随机阵.

可见, 双随机阵是元素皆非负、各行之和及各列之和皆为 1 的方阵. 若记 $\mathbf{1}=(1,\cdots,1)'$, 则 (8.1.5) 和 (8.1.6) 分别为

$$P\mathbf{1}=\mathbf{1},\quad P'\mathbf{1}=\mathbf{1}.$$

相应地, 对次双随机阵, 有 $P\mathbf{1}\preceq\mathbf{1}$, $P'\mathbf{1}\preceq\mathbf{1}$. (此处及以后, 两个向量 $a,b\in R^n$, 记号 $a\leqslant b$ 表示 $a_i\leqslant b_i$, $i=1,\cdots,n$).

下面两个定理刻画了双随机阵的基本性质.

定理 8.1.1 (1) 置换阵是双随机阵;

(2) 双随机阵全体构成一个凸集, 即设 P_1 和 P_2 为任意两个双随机阵, $0\leqslant\alpha\leqslant 1$, 则 $\alpha P_1+(1-\alpha)P_2$ 也是双随机阵.

定理 8.1.2 (1) P_1, P_2 为两个双随机阵, 则它们的 Kronecker 乘积 $P_1\otimes P_2$ 和准对角阵 $\mathrm{diag}(P_1,P_2)$ 也是双随机阵;

(2) 若 P_1 与 P_2 可乘, 则 P_1P_2 也是双随机阵;

(3) 设 $Q=(q_{ij})$ 为正交阵, 则它的 Hadamard 乘积 $P=Q\circ Q=(q_{ij}^2)$ 是双随机阵 (称为正交随机阵). 类似地, 若 $U=(u_{ij})$ 为酉阵, 则 $P=U\circ\overline{U}=(|u_{ij}|^2)$ 也是双随机阵 (称为酉随机阵).

这两个定理的证明都不难, 留给读者作练习.

为了应用双随机阵表征向量的受控关系, 我们还需要 T- 变换的概念.

一个线性变换 T 的矩阵如果具有形式

$$T=\lambda I_n+(1-\lambda)Q,\quad \lambda\in[0,1],$$

则称该变换为 T- 变换. 其中 Q 为恰好对换两个坐标的置换阵. 因为 T 是两个双随机阵 I_n 和 Q 的凸组合, 所以 T 也是双随机阵. 易见, 当 Q 是对换第 j 和第 k 个坐标的置换阵时 $(j<k)$,

$$\begin{aligned}Tx=&(x_1,\cdots,x_{j-1},\lambda x_j+(1-\lambda)x_k,x_{j+1},\cdots,x_{k-1},\\&\lambda x_k+(1-\lambda)x_j,x_{k+1},\cdots,x_n)'.\end{aligned}$$

定理 8.1.3　设 $x \prec y$, 则存在有限个 T- 变换 $T_1, \cdots, T_k$. 使得

$$x = T_k \cdots T_1 y.$$

证明　若存在置换阵 Π, 使得 $x = \Pi y$, 则定理自然成立. 这是因 Π 可表成若干个简单置换阵 (即仅互换单位阵的两行或列所产生的矩阵) 乘积, 而这些简单置换阵也就是 Q 阵, 故它们也是 T 矩阵.

现设 x 不能由 y 经过置换而得到. 不失一般性, 假设 $x_1 \geqslant \cdots \geqslant x_n$, $y_1 \geqslant \cdots \geqslant y_n$. 设 j 为满足 $x_j < y_j$ 的最大下标, 而 k 为大于 j 且 $x_k > y_k$ 的最小下标, 这样的数偶 j, k 一定存在, 因为对 $x_i \neq y_i$ 的最大下标 i, 必有 $x_i > y_i$. 现在选择 j 和 k, 使得

$$y_j > x_j \geqslant x_k > y_k. \tag{8.1.7}$$

记 $d = \min(y_j - x_j, x_k - y_k)$ 和 $\lambda = 1 - \dfrac{d}{y_i - y_k}$, 显然 $0 < \lambda < 1$. 则向量

$$\tilde{y} = (y_1, \cdots, y_{j-1}, y_j - d, y_{j+1}, \cdots, y_{k-1}, y_k + d, y_{k+1}, \cdots, y_n)' \tag{8.1.8}$$

可表为 $\tilde{y} = \lambda y + (1-\lambda)Qy = (\lambda I + (1-\lambda)Q)y = Ty$, 其中 Q 为交换第 j 和第 k 个坐标的置换阵. 现在我们证明: $x \prec \tilde{y}$. 事实上, 从 $x \prec y$ 及 (8.1.8) 知

$$\sum_{i=1}^{t} \tilde{y}_i = \sum_{i=1}^{t} y_i \geqslant \sum_{i=1}^{t} x_i, \qquad t = 1, \cdots, j-1, \tag{8.1.9}$$

$$\sum_{i=1}^{t} \tilde{y}_i = \sum_{i=1}^{t} y_i \geqslant \sum_{i=1}^{t} x_i, \qquad t = k+1, \cdots, n. \tag{8.1.10}$$

由 d 的定义, 知

$$\tilde{y}_j \geqslant x_j, \qquad \tilde{y}_k \geqslant x_k,$$

同时 $y_i = \tilde{y}_i (i = j+1, \cdots, k-1)$. 于是

$$\sum_{i=1}^{t} \tilde{y}_i \geqslant \sum_{i=1}^{t} x_i, \qquad t = j, \cdots, k, \tag{8.1.11}$$

综合 (8.1.9), (8.1.10) 和 (8.1.11) 知: $x \prec \tilde{y}$.

对任意两个 n 维向量 u 和 v, 我们用 $N(u,v)$ 表示 $u_i \neq v_i$ 的 i 的个数. 因为当 $d = y_j - x_j$ 时 $\tilde{y}_j = x_j$, 而当 $d = x_h - y_k$ 时, $\tilde{y}_k = x_k$, 于是 $N(x, \tilde{y}) \leqslant N(x, y) - 1$. 即对 y 施一个 T- 变换后, 既有 $x \prec \tilde{y} = Ty$, 又使 $\tilde{y}$ 与 x 的对应分量不相等的分量个数减少了 1 个. 故将 y 经有限次 T- 变换之后, 可得到 x. 定理证毕.

下面的定理通过双随机阵刻画了受控关系.

定理 8.1.4 对任意 $x, y \in R^n$, $x \prec y \Longleftrightarrow$ 存在双随机阵 P, 使得 $x = Py$.

证明 必要性. 因为 T 矩阵是双随机阵, 且由定理 8.1.2(2) 知 $T_k \cdots T_1$ 也为双随机阵. 故从定理 8.1.3 立得必要性.

充分性. 由受控的置换不变性, 我们不妨设 $x_1 \geqslant \cdots \geqslant x_n$, $y_1 \geqslant \cdots \geqslant y_n$. 因

$$\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} \sum_{j=1}^{n} p_{ij} y_j = \sum_{j=1}^{n} t_j y_j,$$

其中

$$0 \leqslant t_j = \sum_{i=1}^{k} p_{ij} \leqslant 1, \qquad \sum_{j=1}^{n} t_j = k.$$

于是

$$\begin{aligned}
\sum_{i=1}^{k} x_i - \sum_{i=1}^{k} y_i &= \sum_{i=1}^{n} t_i y_i - \sum_{i=1}^{k} y_i \\
&= \sum_{i=1}^{n} t_i y_i - \sum_{i=1}^{k} y_i + y_k \left(k - \sum_{i=1}^{n} t_i\right) \\
&= \sum_{i=1}^{k} (t_i - 1)(y_i - y_k) + \sum_{i=k+1}^{n} t_i (y_i - y_k) \\
&\leqslant 0.
\end{aligned}$$

另外

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} y_j = \sum_{j=1}^{n} y_j \left(\sum_{i=1}^{n} p_{ij}\right) = \sum_{j=1}^{n} y_j,$$

定理得证.

这个定理提供了判定向量受控关系的简单方法.

例 8.1.3 考虑例 8.1.2. 对第一个受控关系, 我们有

$$\begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \triangleq Pa.$$

易见 P 为双随机阵. 而对第二个受控关系, 若在 $(0, \cdots, 0, 1, 0, \cdots, 0)$ 中, 1 位于第 i 个位置, 记之为 e_i, 再记 $a = (a_1, \cdots, a_n)'$. 则对 $a < e_1$, 有 $a = P_1 e_1$, 其中 P_1 为轮迴

矩阵

$$P_1 = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_2 & a_3 & a_4 & \cdots & a_1 \\ a_3 & a_4 & a_5 & \cdots & a_2 \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \end{pmatrix},$$

它是一个双随机阵. 类似地, $a \prec e_2$, 此时 $a = P_2 e_2$, P_2 是将 P_1 的第 1 和 2 两列互换得到的. 一般 $a \prec e_i, i = 1, \cdots, n$, 对应地 $a = P_i e_i$, P_i 是将 P_1 的第 1 和 i 列互换得到的, 它们都是双随机阵.

例 8.1.4　设 $x \in R^n$, $\overline{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$, 则

$$\overline{x}\mathbf{1} = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} x,$$

故 $\overline{x}\mathbf{1} \prec x$.

推论 8.1.1　设 $x^{(1)}, y^{(1)} \in R^k$, $x^{(2)}, y^{(2)} \in R^m$, 且 $x^{(1)} \prec y^{(1)}$, $x^{(2)} \prec y^{(2)}$ 则

$$\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \prec \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}.$$

证明　依定理 8.1.4, 存在两个双随机阵 P_1 和 P_2, 使得 $x^{(1)} = P_1 y^{(1)}$, $x^{(2)} = P_2 y^{(2)}$. 于是

$$\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix} \triangleq P \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}.$$

由定理 8.1.2(1) 知 P 为双随机阵, 故命题得证.

注 4　这个推论的逆命题不真, 即由

$$\begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \prec \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}$$

不能推出 $x^{(i)} \prec y^{(i)}$, $i = 1, 2$. 其中 $x^{(1)} = (x_1, \cdots, x_k)'$, $y^{(1)} = (y_1, \cdots, y_k)'$, 但如果假设了 $x_1 \geqslant \cdots \geqslant x_n$, $y_1 \geqslant \cdots \geqslant y_n$, 且 $\sum\limits_{i=1}^{k} x_i = \sum\limits_{i=1}^{k} y_i$, 则有 $x^{(1)} \prec y^{(1)}$, $x^{(2)} \prec y^{(2)}$.

推论 8.1.2 给定 $y \in R^n$, 则所有受控于 y 的全体 n 维向量构成一个凸集.

证明容易从定理 8.1.4 及 8.1.1(2) 推出.

定理 8.1.5 P 为 $n \times n$ 双随机阵 $\Longleftrightarrow$ 对任意的 $x \in R^n$, $Px \prec x$.

证明 必要性. 设对双随机阵 P, $y = Px$, 则由定理 8.1.4 有 $y \prec x$, 也就是 $Px \prec x$.

充分性. 首先注意到, 对任一 $u \in R^n$, 若有 $u \prec \mathbf{1}_n$, 则必有 $u = \mathbf{1}_n$. 假设对一切 $x \in R^n$, 有 $Px \prec x$, 特别对 $x = \mathbf{1}_n$, 有 $P\mathbf{1}_n \prec \mathbf{1}_n$, 于是 $P\mathbf{1}_n = \mathbf{1}_n$. 再取 $x = e_i$, 即第 i 个分量为 1 其余分量皆为零的 n 维向量, 由 $Pe_i \prec e_i$ 得 $(p_{1i}, \cdots, p_{ni})' \prec e_i$, 于是 $\sum^{n} p_{ji} = 1(i = 1, \cdots, x)$. 即 $P'\mathbf{1} = \mathbf{1}$. 另外, 对任意两个向量 $a \prec b$, 知 $\min_i a_i \geqslant \min_i b_i$, 可推得 $p_{ij} \geqslant 0$, 对一切 i, j 成立. 这就证明了 P 为双随机阵. 证毕.

从前面的讨论我们看到, 双随机阵在受控关系研究中起着非常重要的作用. 对应地, 次双随机阵在弱受控讨论中将起着完全类似的作用. 在进入弱受控讨论之前, 我们先叙述次双随机阵的一些性质, 它们的证明都不难, 留给读者作练习.

定理 8.1.6 (1) 次双随机阵的全体构成一个凸集;

(2) 设 P_1 和 P_2 为次双随机阵, 则 $P_1 \otimes P_2$, $P_1 \circ P_2$ 和 $\mathrm{diag}(P_1, P_2)$ 都是次双随机阵;

(3) 若 P_1, P_2 为同阶次双随机阵, 则 P_1P_2 也是次双随机阵.

定理 8.1.7 设 $P = (p_{ij})$ 为 $n \times n$ 次双随机阵, 则存在一个双随机阵 $D = (d_{ij})$, 满足

$$p_{ij} \leqslant d_{ij}, \ i, \ j = 1, \cdots, n.$$

证明 设 P 是次双随机阵, 但不是双随机阵, 则 P 至少有一行元素之和小于 1, 同时也至少有一列元素之和小于 1. 设第 i 行和第 j 列分别是这样的一个行和列. 定义

$$\varepsilon = \min\left(1 - \sum_{k=1}^{n} p_{ik}, 1 - \sum_{k=1}^{n} p_{kj}\right),$$

$$Q = (q_{ij}) = P + \varepsilon E_{ij},$$

其中 E_{ij} 为 (i, j) 元为 1 其余元素皆为零的 $n \times n$ 阵, 易见 $q_{lk} \geqslant p_{lk}$, $l, k = 1, \cdots, n$, 且 Q 的第 i 列或第 j 列的元素之和至少有一个为 1, 同时 Q 或为次双随机阵或为双随机阵. 如果是前者, 重复上述过程有限次, 便得到满足条件的双随机阵 D. 证毕.

类似于定理 8.1.4, 我们有

定理 8.1.8 设 $x, y \in R_+^n$, 则 $x \prec_w y \Longleftrightarrow$ 存在次双随机阵 P, 使得 $x = Py$. 其中

$$R_+^n = \{x\colon x_i \geqslant 0, \ i = 1, \cdots, n\}. \tag{8.1.12}$$

证明　必要性. 设 $x \prec_w y$. 如果 $x=0$, 则取 $P=0$ 即可. 若 $x \neq 0$, 设 ε 为 x 的最小非零分量, k 为最小正整数, 使得

$$\varepsilon k \geqslant \delta = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i.$$

定义

$$\tilde{x} = \left(x_1, \cdots, x_n, \frac{\delta}{k}, \cdots, \frac{\delta}{k}\right)' \in R^{n+k},$$

$$\tilde{y} = (y_1, \cdots, y_n, 0, \cdots, 0)' \in R^{n+k},$$

显然, $\tilde{x} \prec \tilde{y}$. 由定理 8.1.4 知, 存在一个双随机阵 $\tilde{P}$, 使得 $\tilde{x} = \tilde{P}\tilde{y}$. 记 $\tilde{P}$ 的前 n 行和前 n 列组成的主子阵为 P. 则 P 为次双随机阵, 且 $x = Py$.

充分性. 设 P 为次双随机阵, $x, y \in R_+^n$, 使 $x = Py$. 不妨设 $x_1 \geqslant \cdots \geqslant x_n$, $y_1 \geqslant \cdots \geqslant y_n$. 则

$$\sum_{i=1}^{k} x_j = \sum_{i=1}^{k}\sum_{i=1}^{n} p_{ji} y_i = \sum_{i=1}^{n} t_i y_i,$$

这里

$$0 \leqslant t_i = \sum_{j=1}^{k} p_{ji} \leqslant 1, \quad \sum_{i=1}^{n} t_i \geqslant k.$$

于是

$$\begin{aligned}
\sum_{j=1}^{k} x_j - \sum_{j=1}^{k} y_j &= \sum_{i=1}^{n} y_i t_i - \sum_{i=1}^{k} y_i \\
&\leqslant \sum_{i=1}^{n} y_i t_i - \sum_{i=1}^{k} y_i + y_k\left(k - \sum_{i=1}^{n} t_i\right) \\
&= \sum_{i=1}^{k} (y_i - y_k)(t_i - 1) + \sum_{i=k+1}^{n} t_i (y_i - y_k) \\
&\leqslant 0,
\end{aligned}$$

即 $x \prec_w y$. 证毕.

注 5　定理 8.1.8 中的条件 $x, y \in R_+^n$ 不可少. 例如, 仅有 $x = Py$, P 为次双随机阵, 并不能推出 $x \prec_w y$, 一个简单例子为

$$\begin{pmatrix} 0 \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

类似于定理 8.1.5, 我们可以用弱受控来表征次双随机阵.

定理 8.1.9 P 为 $n\times n$ 次双随机阵 $\Longleftrightarrow$ 对任意 $x\in R_+^n$, 有 $Px\in R_+^n$, 且 $Px\prec_w x$.

证明 设 P 为次双随机阵, $x\in R_+^n$, 显然 $y=Px\in R_+^n$. 依定理 8.1.8 知, $y=Px\prec_w x$. 必要性得证.

充分性. 假设对一切 $x\in R_+^n$, P 满足

$$Px\in R_+^n, \tag{8.1.13}$$

$$Px\prec_w x. \tag{8.1.14}$$

现特别取 $x=\mathbf{1}_n$, 则 (8.1.14) 变为

$$\left(\sum_{j=1}^n p_{1j},\cdots,\sum_{j=1}^n p_{nj}\right)'\prec_w (1,\cdots,1)',$$

这意味着 $\max\limits_i\sum\limits_{j=1}^n p_{ij}\leqslant 1$. 于是 $P\mathbf{1}_n\leqslant\mathbf{1}_n$.

另一方面, 若取 $x=e_j$, 这里 e_j 表示第 j 个分量为 1 其余分量皆为零的 n 维向量. 则 (8.1.14) 变为

$$(p_{1j},p_{2j},\cdots,p_{nj})'\prec_w e_j,\qquad j=1,\cdots,n$$

故有 $\sum\limits_{i=1}^n p_{ij}\leqslant 1, j=1,\cdots,n$, 即 $P'\mathbf{1}_n\leqslant\mathbf{1}_n$.

最后, 在 (8.1.13) 中取 $x=e_j, j=1,\cdots,n$, 可知 $p_{ij}\geqslant 0$, $i,j=1,\cdots,n$. 综上知 P 为次双随机阵. 定理证毕.

在结束这一节的时候, 我们给出受控的若干充分条件, 它们的验证十分容易.

定理 8.1.10 设 $x,y\in R^n$, $x_1\geqslant\cdots\geqslant x_n$, $\sum\limits_{i=1}^n x_i=\sum\limits_{i=1}^n y_i$. 若下列条件之一成立, 必有 $x\prec y$.

(1) 存在 k, $1\leqslant k\leqslant n$, 使得 $x_i\leqslant y_i, i=1,\cdots,k$; $x_i\geqslant y_i$, 当 $i=k+1,\cdots,n$;

(2) y_i-x_i 是关于 i $(i=1,\cdots,n)$ 单调减;

(3) 设 $x_i>0, i=1,\cdots,n$. y_i/x_i 关于 i $(i=1,\cdots,n)$ 是单调减.

证明 设 (1) 成立, 显然 $\sum\limits_{i=1}^j x_i\leqslant\sum\limits_{i=1}^j y_i$, $j\leqslant k$, 同时 $\sum\limits_{i=j}^n x_i\geqslant\sum\limits_{i=j}^n y_i$ $(j=k+1,\cdots,n)$. 但因 $\sum\limits_{i=1}^n x_i=\sum\limits_{i=1}^n y_i$, 故从最后 $n-k$ 个不等式可得

$$\sum_{i=1}^{j-1}x_i\leqslant\sum_{i=1}^{j-1}y_i,\qquad j=k+1,\cdots,n.$$

再结合假设 $x_1 \geqslant \cdots \geqslant x_n$, 便证明了 $x \prec y$.

其次, 我们证明 (2)$\Longrightarrow$(1). 因 $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, 于是 $\sum_{i=1}^{n}(y_i - x_i) = 0$. 由 $y_i - x_i$ 关于 i 的单调减的假设, 知存在 k $(1 \leqslant k \leqslant n)$. 使得 $y_i - x_i \geqslant 0$, $i = 1, \cdots, k$; $y_i - x_i \leqslant 0$, $i = k+1, \cdots, n$, 这就是 (1). 用类似的方法可以证明 (3)$\Longrightarrow$(1). 定理证毕.

定理 8.1.11　设 $x_1 y \in R^n$, $x_1 \geqslant \cdots \geqslant x_n$, $\sum_{i=1}^{n} x_i \leqslant \sum_{i=1}^{n} y_i$. 则定理 8.1.10 中任一条件成立, 都有 $x \prec_w y$.

关于受控关系的进一步性质, 需要应用 Schur 函数的概念, 留在下一节讨论.

§8.2　Schur 函数

Schur 函数是导出不等式的一个重要工具. 本节引进 Schur 函数的定义, 讨论它的基本性质及判定方法, 并给出在向量受控中的一些简单应用.

定义 8.2.1　设 $\boldsymbol{\Phi}(x)$ 为定义在 $A \subset R^n$ 上的实函数, 若对 $x, y \in A$, $x \prec y$, 有 $\boldsymbol{\Phi}(x) \leqslant \boldsymbol{\Phi}(y)$, 则称 $\boldsymbol{\Phi}(x)$ 为 A 上的 Schur 凸函数. 如果进一步, 当 $x \prec y$ 且 x 不是 y 的置换时, 有 $\boldsymbol{\Phi}(x) \prec \boldsymbol{\Phi}(y)$, 则称 $\boldsymbol{\Phi}(x)$ 为 A 上的严格 Schur 凸函数. 当 $A = R^n$ 时, 称 $\boldsymbol{\Phi}(x)$ 分别为 Schur 凸函数和严格 Schur 凸函数.

从定义可以看出, 把 $\boldsymbol{\Phi}(x)$ 称为 Schur 增函数和严格 Schur 增函数更为恰当些.

如果 $-\boldsymbol{\Phi}(x)$ 是 A 上的 Schur 凸函数或严格 Schur 凸函数, 则称 $\boldsymbol{\Phi}(x)$ 为 A 上的 Schur 凹函数或严格 Schur 凹函数. Schur 凸函数和 Schur 凹函数统称为 Schur 函数.

定义 8.2.2　设 $A \subset R^n$, 若对任意 $x \in A$, 对一切置换阵 Π, 有 $\Pi x \in A$, 则称 A 为对称集.

定义 8.2.3　在对称集 A 上定义的函数 $\boldsymbol{\Phi}(x)$, 若对任意的置换阵 Π, 有 $\boldsymbol{\Phi}(x) = \boldsymbol{\Phi}(\Pi x)$, 则称 $\boldsymbol{\Phi}(x)$ 为 A 上的对称函数.

因为对一切 $x \in R^n$ 和任意置换阵 Π, 关系

$$\Pi x \prec x \prec \Pi x, \tag{8.2.1}$$

总成立, 于是我们有

定理 8.2.1　设 $\boldsymbol{\Phi}(x)$ 为对称集 A 上的 Schur 函数, 则 $\boldsymbol{\Phi}(x)$ 必为对称函数.

记

$$D = \{x \in R^n : x_1 \geqslant \cdots \geqslant x_n\}. \tag{8.2.2}$$

定理 8.2.2　设 $\boldsymbol{\Phi}(x)$ 为对称集 A 上的对称函数, 则 $\boldsymbol{\Phi}(x)$ 是 A 上 Schur 凸 (凹) 函数 $\Longleftrightarrow \boldsymbol{\Phi}(x)$ 为 $D \cap A$ 上的 Schur 凸 (凹) 函数.

证明 必要性是显然的, 下证充分性. 设 $x, y \in A, x \prec y$. 记 $\tilde{x} = (x_{(1)}, \cdots, x_{(n)})'$, $x_{(1)} \geqslant \cdots \geqslant x_{(n)}$, 类似地定义 $\tilde{y}$. 因 A 为对称集, 所以 $\tilde{x}, \tilde{y} \in D \cap A$, 故 $\boldsymbol{\Phi}(\tilde{x}) \leqslant \boldsymbol{\Phi}(\tilde{y})$. 但由 $\boldsymbol{\Phi}(x)$ 的对称性, 有 $\boldsymbol{\Phi}(\tilde{x}) = \boldsymbol{\Phi}(x)$, $\boldsymbol{\Phi}(\tilde{y}) = \boldsymbol{\Phi}(y)$. 于是 $\boldsymbol{\Phi}(x) \leqslant \boldsymbol{\Phi}(y)$, 即 $\boldsymbol{\Phi}(x)$ 为 Schur 凸函数. 证毕.

这个定理表明, 当我们研究一个对称集上对称函数的 Schur 凸 (凹) 性时, 只需考察它在 D 上的 Schur 凸 (凹) 性. 这一性质会给我们带来很大方便, 后面将多次用到.

Schur 函数是一个多元函数, 直接用定义判定它是很不方便的. 下面我们建立一些较简便的判定方法.

定理 8.2.3 $\boldsymbol{\Phi}(x)$ 为 D 上 Schur 凸 (严格 Schur 凸) 函数 $\Longleftrightarrow$ 对任意的 $z \in D$, 函数 $\boldsymbol{\Phi}(z_1, \cdots, z_{k-1}, z_k + \varepsilon, z_{k+1} - \varepsilon, z_{k+2}, \cdots, z_n)$ 是 ε 的增 (严格增) 函数, 其中 ε 的变化范围为

$$
\begin{aligned}
&0 \leqslant \varepsilon \leqslant z_2 - z_3, \ \text{当 } k = 1, \\
&0 \leqslant \varepsilon \leqslant \min\{z_{k-1} - z_k, z_{k+1} - z_{k+2}\}, \ \text{当 } 2 \leqslant k \leqslant n-2, \qquad (8.2.3)\\
&0 \leqslant \varepsilon \leqslant z_{n-2} - z_{n-1}, \ \text{当 } k = n-1.
\end{aligned}
$$

证明 记

$$
S = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}_{n \times n},
$$

$$
\tilde{x} = (\tilde{x}_1, \cdots, \tilde{x}_n)' = Sx,
$$

$$
\tilde{y} = (\tilde{y}_1, \cdots, \tilde{y}_n)' = Sy.
$$

显然, $x \prec y \Longleftrightarrow \tilde{x}_i \leqslant \tilde{y}_i, i = 1, \cdots, n-1, \tilde{x}_n = \tilde{y}_n$. 于是事实 "$x \prec y$, 有 $\boldsymbol{\Phi}(x) \leqslant \boldsymbol{\Phi}(y)$" 等价于 "对 $\tilde{x}_i \leqslant \tilde{y}_i, i = 1, \cdots, n-1, \tilde{x}_n = \tilde{y}_n$, 有 $\boldsymbol{\Phi}(\tilde{x}_1, \tilde{x}_2 - \tilde{x}_1, \cdots, \tilde{x}_n - \tilde{x}_{n-1}) \leqslant \boldsymbol{\Phi}(\tilde{y}_1, \tilde{y}_2 - \tilde{y}_1, \cdots, \tilde{y}_n - \tilde{y}_{n-1})$". 后者表明 $\boldsymbol{\Phi}(\tilde{z}_1, \tilde{z}_2 - \tilde{z}_1, \cdots, \tilde{z}_n - \tilde{z}_{n-1})$ 对每个变量 $\tilde{z}_i$ 是增函数 (当其他变量 $\tilde{z}_j$ $(j \neq i)$ 固定时), 这又等价于 $\boldsymbol{\Phi}(z_1, \cdots, z_{k-1}, z_k + \varepsilon, z_{k+1} - \varepsilon, z_{k+2}, \cdots, z_n)$ 是 ε 的增函数. ε 的变化范围 (8.2.3) 是为了 $\boldsymbol{\Phi}$ 有定义. 证毕.

在微积分中, 我们曾经用函数的一阶导数的正负号来判定函数的单调性. 对现在的情形, 我们也有相类似的结果.

记

$$
\frac{\partial \boldsymbol{\Phi}}{\partial x} = \left(\frac{\partial \boldsymbol{\Phi}}{\partial x_1}, \cdots, \frac{\partial \boldsymbol{\Phi}}{\partial x_n}\right),
$$

$$
D^\circ = \{x \in R^n, \ x_1 > x_2 > \cdots > x_n\},
$$

易见 D° 是 D 的内部.

定理 8.2.4 设 $\boldsymbol{\Phi}(x)$ 为 D 上连续函数, 且在 D° 上连续可微. 则 $\boldsymbol{\Phi}(x)$ 在 D 上为 Schur 凸函数 $\Longleftrightarrow \dfrac{\partial \boldsymbol{\Phi}(z)}{\partial z} \in D$, 对一切 $y \in D^\circ$. 即

$$\frac{\partial \boldsymbol{\Phi}(z)}{\partial z_1} \geqslant \frac{\partial \boldsymbol{\Phi}(z)}{\partial z_2} \geqslant \cdots \geqslant \frac{\partial \boldsymbol{\Phi}(z)}{\partial z_n}, \text{ 对一切 } z \in D^\circ \text{ 成立.}$$

证明 首先注意到, 我们只需要在 D° 上证明定理. 事实上, 设 $x, y \in D$. 则对任意 $\varepsilon > 0$, 定义

$$x_\varepsilon = (x_1 + n\varepsilon,\ x_2 + (n-1)\varepsilon, \cdots, x_n + \varepsilon)' \in D^\circ,$$

且当 $\varepsilon \to 0$ 时, $x_\varepsilon \to x$. 类似地定义 $y_\varepsilon \in D^\circ$, 当 $\varepsilon \to 0$ 时, $y_\varepsilon \to y$. 容易看到由 $x \prec y$ 可推出 $x_\varepsilon \prec y_\varepsilon$. 若 $\boldsymbol{\Phi}$ 是 D° 上的 Schur 凸函数, 则 $\boldsymbol{\Phi}(x_\varepsilon) \leqslant \boldsymbol{\Phi}(y_\varepsilon)$, 对一切 $\varepsilon > 0$ 成立, 根据 $\boldsymbol{\Phi}$ 在 D 上的连续性, 有

$$\boldsymbol{\Phi}(x) = \lim_{\varepsilon \to 0} \boldsymbol{\Phi}(x_\varepsilon) \leqslant \lim_{\varepsilon \to 0} \boldsymbol{\Phi}(y_\varepsilon) = \boldsymbol{\Phi}(y).$$

这就证明了 D° 上的 Schur 凸函数一定是 D 上 Schur 凸函数

根据定理 8.2.3, $\boldsymbol{\Phi}$ 是 D° 上 Schur 凸函数, 当且仅当对 $z \in D^\circ$, 函数 $\boldsymbol{\Phi}(z_1, \cdots, z_{k-1}, z_k + \varepsilon, z_{k+1} - \varepsilon, z_{k+2}, \cdots, z_n)$ 在 (8.2.3) 定义的 ε 区域上是 ε 的增函数. 这又等价于

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \boldsymbol{\Phi}(z_1, \cdots, z_{k-1}, z_k + \varepsilon, z_{k+1} - \varepsilon, z_{k+2}, \cdots, z_n) \geqslant 0, \quad z \in D^\circ.$$

即

$$\frac{\partial \boldsymbol{\Phi}}{\partial z_k} - \frac{\partial \boldsymbol{\Phi}}{\partial z_{k+1}} \geqslant 0, \quad z \in D^\circ.$$

定理证毕.

在微积分中, 为了判定函数严格增, 对一阶导数等于零的点, 我们要看它的二阶导数的符号. 关于这一点有如下事实: 设 $f(x)$ 为定义在 $[a, b]$ 上的实函数, $f'(x) \geqslant 0$. 且对一切使 $f'(x) = 0$ 的 x, 有 $f''(x) > 0$. 则 $f(x)$ 为严格增函数. 应用此事实及定理 8.2.3 我们得到

定理 8.2.5 设 $\boldsymbol{\Phi}$ 为 D 上连续实函数, 在 D° 上二次可微. $\boldsymbol{\Phi}$ 为 D 上 Schur 凸函数, 且对满足

$$\frac{\partial \boldsymbol{\Phi}}{\partial z_k} = \frac{\partial \boldsymbol{\Phi}}{\partial z_{k+1}}$$

的 z, 有

$$\frac{\partial^2 \boldsymbol{\Phi}}{\partial z_k \partial z_k} - \frac{\partial^2 \boldsymbol{\Phi}}{\partial z_k \partial z_{k+1}} - \frac{\partial^2 \boldsymbol{\Phi}}{\partial z_{k+1} \partial z_k} + \frac{\partial^2 \boldsymbol{\Phi}}{\partial z_{k+1} \partial z_{k+1}} > 0,$$

则 $\boldsymbol{\Phi}$ 为 D 上严格 Schur 凸函数.

设 $I \subset R$ 为一开区间, 易见 $I^n = \{x \in R^n, x_i \in I, i = 1, \cdots, n\}$ 为一对称集. 应用定理 8.2.1, 8.2.2 和 8.2.4, 我们立即得到如下定理.

定理 8.2.6 设 $\boldsymbol{\Phi}$ 为 $I^n \to R$ 的连续可微函数, 则 $\boldsymbol{\Phi}$ 在 I^n 上为 Schur 凸函数当且仅当

(1) $\boldsymbol{\Phi}$ 为 I^n 上对称函数;

(2) $\dfrac{\partial \boldsymbol{\Phi}}{\partial z_1} \geqslant \cdots \geqslant \dfrac{\partial \boldsymbol{\Phi}}{\partial z_n}$, 对一切 $z \in D \cap I^n$.

注 1 借助于$\boldsymbol{\Phi}$的对称性, 上面的条件 (2) 可以改为: 对任意$i \neq j$及一切$z \in I^n$, 有

$$(z_i - z_j)\left(\frac{\partial \boldsymbol{\Phi}}{\partial z_i} - \frac{\partial \boldsymbol{\Phi}}{\partial z_j}\right) \geqslant 0. \tag{8.2.4}$$

同样, 由 $\boldsymbol{\Phi}$ 的对称性, (8.2.4) 可进一步改为

$$(z_1 - z_2)\left(\frac{\partial \boldsymbol{\Phi}}{\partial z_1} - \frac{\partial \boldsymbol{\Phi}}{\partial z_2}\right) \geqslant 0. \tag{8.2.5}$$

条件 (8.2.4) 和 (8.2.5) 都称为 Schur 条件. 它们在应用上非常方便. 事实上, 许多 Schur 凸函数都是用这些条件证明的.

在 Schur 函数理论中, 下面的定理占有十分重要的地位. 它表明, 当我们要证明 $\boldsymbol{\Phi}(x)$ 为 Schur 凸函数时, 只需考虑 $n = 2$ 的情况. 即把判定一个 n 元函数为 Schur 凸函数的问题归结为二元函数的对应问题, 后者显然容易得多.

定理 8.2.7 设 $\boldsymbol{\Phi}$ 为 R^n 上对称函数. 固定 $z_3 = z_3^0, \cdots, z_n = z_n^0$, 若函数

$$\boldsymbol{\Phi}(z) \triangleq \boldsymbol{\Phi}(z_1, z_2, z_3^0, \cdots, z_n^0)$$

为 R^2 上的 Schur 凸函数, 则 $\boldsymbol{\Phi}$ 为 R^n 上 Schur 凸函数.

证明 设 $x \prec y$, 由定理 8.1.3, 存在有限个 T- 变换 $T_1, \cdots, T_k$, 使得

$$x = T_k \cdots T_1 y.$$

记 $\tilde{y} = T_1 y$. 由 T- 变换的结构知, $\tilde{y}$ 与 y 至多有两个对应分量不相等. 由 $\boldsymbol{\Phi}$ 的对称性, 不妨假设它们是前两个分量, 于是 $\tilde{y}_i = y_i$, $i \geqslant 3$. 因为 $\tilde{y} \prec y$, 所以 $(\tilde{y}_1, \tilde{y}_2)' \prec (y_1, y_2)'$. 因而

$$\boldsymbol{\Phi}(\tilde{y}) = \boldsymbol{\Phi}(\tilde{y}_1, \tilde{y}_2, y_3^0, \cdots, y_n^0) \leqslant \boldsymbol{\Phi}(y_1, y_2, y_3^0, \cdots, y_n^0) = \boldsymbol{\Phi}(y),$$

即 $\boldsymbol{\Phi}(T_1 y) \leqslant \boldsymbol{\Phi}(y)$. 于是我们有

$$\begin{aligned}\boldsymbol{\Phi}(x) = \boldsymbol{\Phi}(T_k \cdots T_1 y) &\leqslant \boldsymbol{\Phi}(T_{k-1} \cdots T_1 y) \leqslant \cdots \leqslant \boldsymbol{\Phi}(T_1 y) \\ &\leqslant \boldsymbol{\Phi}(y).\end{aligned}$$

证毕.

现在我们应用定理 8.2.7 证明如下重要事实.

定理 8.2.8 设 $\boldsymbol{\Phi}$ 是 R^n 上对称凸 (严格凸) 函数, 则 $\boldsymbol{\Phi}$ 是 R^n 上 (严格)Schur 凸函数.

证明 根据定理 8.2.7, 我们只需对 $n=2$ 进行证明. 设 $x<y$, 则存在 $\alpha\in[0,1]$, 使得

$$x_1=\alpha y_1+(1-\alpha)y_2,$$

$$x_2=(1-\alpha)y_1+\alpha y_2.$$

由 $\boldsymbol{\Phi}$ 的凸性及对称性, 有

$$\begin{aligned}\boldsymbol{\Phi}(x)=\boldsymbol{\Phi}(x_1,x_2)&=\boldsymbol{\Phi}(\alpha y_1+(1-\alpha)y_2,\ (1-\alpha)y_1+\alpha y_2)\\&=\boldsymbol{\Phi}\left(\alpha\begin{pmatrix}y_1\\y_2\end{pmatrix}+(1-a)\begin{pmatrix}y_2\\y_1\end{pmatrix}\right)\\&\leqslant\alpha\boldsymbol{\Phi}(y_1,y_2)+(1-\alpha)\boldsymbol{\Phi}(y_2,y_1)\qquad(8.2.6)\\&=\alpha\boldsymbol{\Phi}(y_1,y_2)+(1-\alpha)\boldsymbol{\Phi}(y_1,y_2)\\&=\boldsymbol{\Phi}(y_1,y_2)\\&=\boldsymbol{\Phi}(y),\end{aligned}$$

这就证明了 $\boldsymbol{\Phi}$ 为 R^n 上的 Schur 凸函数.

设 $\boldsymbol{\Phi}$ 为严凸函数, 则 $\boldsymbol{\Phi}(x)=\boldsymbol{\Phi}(y)\Longleftrightarrow$(8.2.6) 式取等号 $\Longleftrightarrow\alpha=0$ 或 $1\Longleftrightarrow x_1=y_1$, $x_2=y_2$ 或 $x_1=y_2,x_2=y_1$, 即 y 为 x 的置换. 这就证明了 $\boldsymbol{\Phi}$ 为严格 Schur 凸函数. 定理证毕.

结合定理 8.2.7, 我们立即有

推论 8.2.1 当任意 $n-2$ 个自变量固定时, $\boldsymbol{\Phi}$ 为剩余两个自变量的对称凸函数, 则 $\boldsymbol{\Phi}$ 为 R^n 上的 Schur 凸函数.

现在我们讨论一些非常有用的特殊形式的函数的 Schur 凸 (凹) 性. 它们是如下两类复合函数

(1) $\psi(x)=h(\varphi_1(x),\cdots,\varphi_k(x))$,

其中, $\varphi_1(x),\cdots,\varphi_k(x)$ 为定义在 $A\subset R^n$ 上的实函数, h 为定义在 R^k 上实函数;

(2) $\psi(x)=f(g(x_1),\cdots,g(x_n))$,

其中, g 为定义在 R 上的实函数, f 为定义在 R^n 上的实函数.

我们的问题是, 当函数 $f,g,h,\varphi_i,\cdots,\varphi_k$ 满足哪些条件时, 复合函数 ψ 为 Schur 函数.

定理 8.2.9 设 h 是增 (减) 函数 (即固定任意 $k-1$ 个自变量, h 是另一自变量的增 (减) 函数), 则

(1) 当 $\varphi_i(x)$ 均为 Schur 凸函数时, $\psi(x)$ 为 Schur 凸 (凹) 函数;

(2) 当 $\varphi_i(x)$ 均为 Schur 凹函数时, $\psi(x)$ 为 Schur 凹 (凸) 函数.

证明 (1) 设 $x \prec y$, 由 $\varphi_i(x)$ 是 Schur 凸性得

$$\varphi_i(x) \leqslant \varphi_i(y), \qquad i=1,\cdots,k,$$

再由 h 对每个自变元为增函数, 于是

$$\begin{aligned}\psi(x) &= h(\varphi_1(x),\cdots,\varphi_k(x)) \leqslant h(\varphi_1(y),\cdots,\varphi_k(y))\\ &= \psi(y),\end{aligned}$$

这就证明了 ψ 为 Schur 凸函数.

(2) 证明类似于 (1), 从略. 定理证毕.

推论 8.2.2 (1) 若干个 Schur 凸函数的非负线性组合仍为 Schur 函数, 于是 Schur 凸函数构成一凸锥;

(2) 一个 Schur 凸函数的增函数仍为 Schur 凸函数.

证明 (1) 在定理中取 $\varphi_i(x)=x_i$, $h(x_1,\cdots,x_k)=\sum\limits_{i=1}^k \alpha_i x_i$, $\alpha_i \geqslant 0$, 即得欲证.

(2) 显然. 证毕.

例 8.2.1 设 $\varphi_1,\cdots,\varphi_k$ 为 Schur 凸 (凹) 函数, 则

$$\min(\varphi_1,\cdots,\varphi_k) \text{ 和 } \max(\varphi_1,\cdots,\varphi_k)$$

均为 Schur 凸 (凹) 函数.

例 8.2.2 设 $\varphi_1,\cdots,\varphi_k$ 为 Schur 凸 (凹) 函数, 且对一切 i 及 x, 有 $\varphi_i(x) \geqslant 0$. 则

$$\psi(x)=\prod_{i=1}^k \varphi_i(x)$$

为 Schur 凸 (凹) 函数.

现在我们转入第二类复合函数的 Schur 凸性的讨论.

定理 8.2.10 设 g 为凸 (凹) 函数, f 为增 (减)Schur 凸函数, 则 $\psi=f(g(x_1),\cdots,g(x_n))$ 为 Schur 凸函数.

证明 首先由定理 8.2.1 知, f 必为对称函数, 再依定理 8.2.7, 我们只需对 $n=2$ 证明定理.

当 $n=2$ 时, 由 $x \prec y$, 必存在 $\alpha \in [0,1]$, 使得

$$x_1 = \alpha y_1 + (1-\alpha) y_2,$$

$$x_2 = (1-\alpha) y_1 + \alpha y_2.$$

由 g 的凸性及 f 是增函数, 我们得

$$\begin{aligned}\psi(x) &= f(g(x_1), g(x_2)) \\ &= f(g(\alpha y_1 + (1-\alpha) y_2), g((1-\alpha) y_1 + \alpha y_2)) \\ &\leqslant f(\alpha g(y_1) + (1-\alpha) g(y_2), (1-\alpha) g(y_1) + \alpha g(y_2)) \\ &= f\left(\alpha \begin{pmatrix} g(y_1) \\ g(y_2) \end{pmatrix} + (1-\alpha) \begin{pmatrix} g(y_2) \\ g(y_1) \end{pmatrix}\right).\end{aligned} \tag{8.2.7}$$

因为

$$\alpha \begin{pmatrix} g(y_1) \\ g(y_2) \end{pmatrix} + (1-\alpha) \begin{pmatrix} g(y_2) \\ g(y_1) \end{pmatrix} < \begin{pmatrix} g(y_1) \\ g(y_2) \end{pmatrix},$$

于是由 f 的 Schur 凸知

$$\begin{aligned}\psi(x) \leqslant f\left(\alpha \begin{pmatrix} g(y_1) \\ g(y_2) \end{pmatrix} + (1-\alpha) \begin{pmatrix} g(y_2) \\ g(y_1) \end{pmatrix}\right) \leqslant f(g(y_1), g(y_2)) \\ = \psi(y).\end{aligned}$$

定理得证.

推论 8.2.3 设 $I \subset R$ 的一个区间. g 为 $I \to R$ 上凸 (严凸) 函数. 则

$$\psi(x) = \sum_{i=1}^{n} g(x_i) \tag{8.2.8}$$

为 I^n 上 (严格)Schur 凸函数.

证明 在定理 8.2.10 中, 取 $f(x) = \sum_{i=1}^{n} x_i$, 即可得 $\boldsymbol{\Phi}(x)$ 的 Schur 凸性. 当 $g(x)$ 为严凸函数时, 因 $\psi(x) = \psi(y) \Longleftrightarrow (8.2.7)$ 式等号成立 $\Longleftrightarrow \alpha = 0$ 或 $1 \Longleftrightarrow y$ 为 x 的置换. 这就证明了 ψ 的严格凸性. 证毕.

例 8.2.3 在 $x_i > 0$, $i = 1, \cdots, n$ 上, 函数

$$\boldsymbol{\Phi}(x) = \sum_{i=1}^{n} \frac{1}{x_i}$$

为严格 Schur 凸函数.

例 8.2.4 设 $\alpha > 0$. 则

$$\varPhi(x) = \sum_{i=1}^{n} \left(x_i + \frac{1}{x_i}\right)^a$$

在 $0 < x_i \leqslant 1, i = 1, \cdots, n$ 上为严格 Schur 凸函数.

设$x_i > 0, \sum\limits_{i=1}^{n} x_i = 1$, 则$\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)' \prec x$. 利用$\varPhi(x)$的严格凸性, 得$\varPhi\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) < \varPhi(x)$, 即

$$\sum_{i=1}^{n} \left(x_i + \frac{1}{x_i}\right)^a \geqslant \frac{(n^2+1)^a}{n^{a-1}}.$$

例 8.2.5 函数 $\varPhi(x) = \sum\limits_{i=1}^{n} \log x_i$ 在 $x_i > 0, i = 1, \cdots, n$ 上为严格 Schur 凸.

据此及算术平均和几何平均不等式, 我们立即得到下面的不等式. 设 $x_i \geqslant 0$, $i = 1, \cdots, n$, $M(x) = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, $G(x) = (x_1 x_2 \cdots x_n)^{1/n}$. 则

$$\sum_{i=1}^{n} \log(1 + x_i) \geqslant n \log(1 + M(x)) \geqslant n \log(1 + G(x)).$$

等价地

$$\prod_{i=1}^{n} (1 + x_i) \geqslant (1 + G(x))^n.$$

推论 8.2.4 设 $I \subset R$ 为一区间, g 为 I 上连续的正值函数, 且 $\log g$ 为 (严格) 凸函数, 则

$$\varPhi(x) = \prod_{i=1}^{n} g(x_i) \tag{8.2.9}$$

为 (严格)Schur 凸函数.

证明 因 $\log \varPhi(x) = \Sigma \log g(x_i)$. 根据推论 8.2.4, 从 $\log g$ 的凸性可推出 $\log \varPhi(x)$ 的 Schur 凸. 再利用

$$\varPhi(x) = e^{\log \phi(x)},$$

由推论 8.2.2 知, $\varPhi(x)$ 为 Schur 凸函数. 证毕.

例 8.2.6 函数

$$\varPhi_1(x) = \prod_{i=1}^{n} \frac{1+x_i}{x_i}, \quad \varPhi_2(x) = \prod_{i=1}^{n} \frac{1-x_i}{x_i}, \quad \varPhi_3(x) = \prod_{i=1}^{n} \frac{1+x_i}{1-x_i}$$

在 $x_i > 0, i = 1, \cdots, n$ 上都是严格 Schur 凸函数.

容易验证 $\log\dfrac{1+u}{u}$, $\log\dfrac{1-u}{u}$ 和 $\log\dfrac{1+u}{1-u}$ 在 $u>0$ 上都是严格凸函数, 于是由推论 8.2.4 知, $\varPhi_i(x)$, $i=1,2,3$ 都是严格 Schur 凸. 设 $x_i>0$, $i=1,\cdots,n$, $\sum\limits_{i=1}^n x_i=1$. 则有

$$\left(\frac{1}{n},\cdots,\frac{1}{n}\right)\prec x.$$

据此及 $\varPhi_i(x)$, $i=1,2,3$ 的严格 Schur 凸, 立得下面三个不等式

$$\prod_{i=1}^n\frac{1+x_i}{x_i}\geqslant(n+1)^n,$$
$$\prod_{i=1}^n\frac{1-x_i}{x_i}\geqslant(n-1)^n,$$
$$\prod_{i=1}^n\frac{1+x_i}{1-x_i}\geqslant\left(\frac{n+1}{n-1}\right)^n$$

对一切满足 $\sum\limits_{i=1}^n x_i=1$ 的正数 $x_1,\cdots,x_n$ 成立, 且等号成立 $\Longleftrightarrow x_i=\dfrac{1}{n}$, $i=1,\cdots,n$. 前两个不等式是 Klamkin 和 Newman(1970) 证明的, 第三个由 Klamkin(1975) 证明的.

现在我们考虑另外一类函数的 Schur 凹凸性. 记

$$S_0\equiv1,\quad S_1(x)=\sum_{i=1}^n x_i,$$
$$S_2(x)=\sum_{i<j}x_ix_j,\cdots,S_n(x)=\prod_{i=1}^n x_i,$$

称 $S_k(x)$ 为第 k 个初等对称函数.

定理 8.2.11　初等对称函数 $S_k(x)$ 是区域 $x_i\geqslant0$, $i=1,\cdots,n$ 上的增 Schur 凹函数 (即它们是 Schur 凹函数且是每个自变量的增函数). 当 $k\geqslant2$ 时, $S_k(x)$ 为这域 $x_i>0$, $i=1,\cdots,n$ 上的严格 Schur 凹函数

证明　注意到

$$\frac{\partial S_k}{\partial x_i}=S_{k-1}(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n),$$

故当 $x_i\geqslant x_j$ 时,

$$S_{k-1}(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)\leqslant S_{k-1}(x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_n),$$

而当 $x_i\leqslant x_j$ 时, 上式中的不等号反向. 故

$$(x_i-x_j)\left(\frac{\partial S_k}{\partial x_i}-\frac{\partial S_k}{\partial x_j}\right)\leqslant0.$$

由定理 8.2.6 之后的注, 知 S_k 为 Schur 凹函数. 再由定理 8.2.5, 可推得 $k \geqslant 2$ 时的严格 Schur 凹性. 证毕.

例 8.2.7 设 $x, y \in R^n$, $x_i \geqslant 0$, $y_i \geqslant 0$, $i = 1, \cdots, n$, 则当 $x \prec y$ 时, 有

$$\prod_{i=1}^{n} x_i \geqslant \prod_{i=1}^{n} y_i.$$

这是定理 8.2.11 当 $k = n$ 时的结论. 这个不等式称为 Schur 不等式.

下面我们不加证明地叙述初等对称函数的其他一些性质. 其证明见 Marshall 和 Olkin(1979, p.70~80.)

定理 8.2.12 (1) 在 $x_i \geqslant 0, i = 1, \cdots, n$ 上, $\Phi(x) = (S_k(x))^{1/k}$ 是增 Schur 凹函数;

(2) 在 $x_i > 0, i = 1, \cdots, n$ 上, $\Phi(x) = \dfrac{S_k(x)}{S_{k-1}(x)}$ 是 Schur 凹函数;

(3) 对 $1 \leqslant p \leqslant k \leqslant n$, 有

$$G_{k,p}(x) = \left(\frac{S_k(x)}{S_{k-p}(x)}\right)^{1/p}$$

在 $x_i > 0, i = 1, \cdots, n$ 上为 Schur 凹函数.

在结束这一节的时候, 我们证明凸函数的一个重要性质 —— 保序性, 它在下节讨论中将要用到.

定理 8.2.13 对一切定义在 R 上的凸函数 g, 则

$$x \prec y \Longleftrightarrow (g(x_1), \cdots, g(x_n))' \prec_w (g(y_1), \cdots, g(y_n))'.$$

证明 充分性. 只需取两个特殊凸函数 $g_1(u) = u$ 和 $g_2(u) = -u$ 即可.

必要性. 设 $x \prec y$, 由定理 8.1.3 知, 存在 k 个 T- 变换, 使 $x = T_k \cdots T_1 y$. 我们知道, 对任一向量, 每做一次 T- 变换, 最多有两个分量发生改变, 再应用弱受控的传递性, 我们只需对 $n = 2$ 进行证明.

当 $n = 2$ 时, 对任意 $x \prec y$, 必存在 $a \in [0, 1]$, 使得

$$\begin{aligned} x_1 &= \alpha y_1 + (1-\alpha) y_2, \\ x_2 &= (1-\alpha) y_1 + \alpha y_2. \end{aligned} \tag{8.2.10}$$

于是由 g 的凸性, 我们有

$$\begin{aligned} &\max\{g(x_1), g(x_2)\} \\ =&\max\{g(\alpha y_1 + (1-\alpha) y_2), g((1-\alpha) y_1 + \alpha y_2)\} \\ \leqslant&\max\{\alpha g(y_1) + (1-\alpha) g(y_2),\ (1-\alpha) g(y_1) + \alpha g(y_2)\} \\ \leqslant&\max\{g(y_1), g(y_2)\}. \end{aligned}$$

更进一步, 由 g 的凸性及 (8.2.10), 可得

$$g(x_1)+g(x_2)\leqslant g(y_1)+g(y_2),$$

即 $(g(x_1),g(x_2))'\prec_w (g(y_1),g(y_2))'$. 证毕.

例 8.2.8　设 $x,y\in R^n$, $x\prec y$, 则

(1) $(x_1^2,\cdots,x_n^2)'\prec(y_1^2,\cdots,y_n^2)'$;

(2) $(|x_1|,\cdots,|x_n|)'\prec(|y_1|,\cdots,|y_n|)'$.

§8.3　Hermite 阵

我们曾经指出, 受控和 Schur 函数的结合是导出不等式的一个重要工具, 许多常见的不等式都可以用这一方法获得. 本书不打算讨论这一方面的内容, 感兴趣的读者可阅读 Marshall 和 Olkin(1979) 的名著 “Inequalities: Theory of Majorization and its Applications” . 在本章的后面几节, 我们将把受控和 Schur 函数应用于矩阵论中一些不等式的推导. 它们主要是方阵的对角元向量、特征值向量和奇异值向量之间的受控关系. 我们先从 Hermite 阵开始.

定理 8.3.1　设 $A=(a_{ij})$ 为 $n\times n$ Hermite 阵, $a=(a_{11},\cdots,a_{nn})'$ 为 A 的对角元向量, $\lambda=(\lambda_1,\cdots,\lambda_n)'$ 为 A 的特征值向量. 则 $a\prec\lambda$.

证明　因 A 为 Hermite 阵, 故存在酉阵 $U=(u_{ij})$, 使得 $A=U\Lambda U^*$, 其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$(见定理 1.4.1). 注意到

$$a_{ii}=\sum_j u_{ij}\overline{u}_{ij}\lambda_j=\sum_j p_{ij}\lambda_j,$$

其中 $p_{ij}=u_{ij}\overline{u}_{ij}$, 由 U 是酉阵可知 $P=(p_{ij})$ 为双随机阵. 于是 $a=P\lambda$. 利用定理 8.1.4 便有 $a\prec\lambda$. 证毕.

把事实 “$a\prec\lambda$” 与 Schur 函数相结合, 我们可以导出许多关于 Hermite 阵的对角元和特征值的不等式. 下面是最常用的几类.

推论 8.3.1　设 $A=(a_{ij})$ 为 $n\times n$ Hermite 阵. $\lambda_1,\cdots,\lambda_n$ 为其特征值. 则对任意定义在 R 上凸函数 $\boldsymbol{\Phi}$, 有

$$\sum_{i=1}^n\boldsymbol{\Phi}(a_{ii})\leqslant\sum_{i=1}^n\boldsymbol{\Phi}(\lambda_i).$$

证明　由定理 8.3.1 和推论 8.2.3 立得本推论.

推论 8.3.2　设 A 为 $n\times n$ 半正定 Hermite 阵, S_k 为初等对称函数. 则

$$S_k(\lambda_1,\cdots,\lambda_n)\geqslant S_k(a_{11},\cdots,a_{nn}),\quad k=1,\cdots,n. \tag{8.3.1}$$

证明 因为 A 为半正定 Hermite 阵, 故 $a_{ii} \geqslant 0, \lambda_i \geqslant 0, i = 1, \cdots, n$. 应用定理 8.3.1 和定理 8.2.11 立得所要结论.

下面的推论是 §3.3 证明过的 Hadamard 不等式. 此处它们可以作为 (8.3.1) 中 $k = n$ 的特殊情形而导出.

推论 8.3.3(Hadamard 不等式)

(1) 设 A 为 $n \times n$ 半正定 Hermite 阵. 则

$$\det A \leqslant \prod_{i=1}^{n} a_{ii};$$

(2) 设 A 为 $n \times m$ 矩阵, 则

$$\det AA^* \leqslant \prod_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2.$$

证明 在 (8.3.1) 中, $k = n$ 即得 (1), 对 AA^* 应用 (8.3.1) 中 $k = n$, 即得 (2).

推论 8.3.4 设 A 为 $n \times n$ 正定 Hermite 阵, 则对初等对称函数 S_k, 有

$$\begin{aligned}\frac{S_n(a_{11}, \cdots, a_{nn})}{S_n(\lambda_1, \cdots, \lambda_n)} &\geqslant \frac{S_{n-1}(a_{11}, \cdots, a_{nn})}{S_{n-1}(\lambda_1, \cdots, \lambda_n)} \geqslant \cdots \\ &\geqslant \frac{S_1(a_{11}, \cdots, a_{nn})}{S_1(\lambda_1, \cdots, \lambda_n)} = 1.\end{aligned}$$

结论可直接从定理 8.3.1 及定理 8.2.12(2) 推出.

定理 8.3.1 的逆定理也成立. 为了证明这个事实, 我们先证明如下两个引理.

引理 8.3.1 给定两组实数 $c_1, \cdots, c_{n-1}$ 和 $\lambda_1, \cdots, \lambda_n$, 满足

$$\lambda_1 \geqslant c_1 \geqslant \lambda_2 \geqslant \cdots \geqslant c_{n-1} \geqslant \lambda_n.$$

则存在 $n \times n$ 实对称阵

$$A = \begin{pmatrix} \Lambda_c & b \\ b' & d \end{pmatrix}$$

以 $\lambda_1, \cdots, \lambda_n$ 为特征值, 其中 $\Lambda_c = \mathrm{diag}(c_1, \cdots, c_{n-1})$, b 为 $(n-1) \times 1$ 实向量.

证明 应用引理 3.2.1(3) 知, A 的特征方程为

$$\begin{aligned}\det(\lambda I_n - A) &= \det(\lambda I_{n-1} - \Lambda_c)(\lambda - d - b'(\lambda I_{n-1} - \Lambda_c)^{-1} b) \\ &= \prod_{i=1}^{n-1} (\lambda_i - c_i) \left(\lambda - d - \sum_{i=1}^{n-1} \frac{b_i^2}{\lambda_i - c_i} \right) = 0. \qquad (8.3.2)\end{aligned}$$

因为 $\lambda_1, \cdots, \lambda_n$ 和 $c_1, \cdots, c_{n-1}$ 都是给定的, 所以我们的问题是选择 $b = (b_1, \cdots, b_{n-1})'$ 和 d, 使得 $\lambda_1, \cdots, \lambda_n$ 为 (8.3.2) 的根.

记

$$f(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i), \quad g(\lambda) = \prod_{i=1}^{n-1} (\lambda - c_i).$$

先设 $c_1, \cdots, c_{n-1}$ 互异, 则有

$$f(\lambda) - g(\lambda) \left[\lambda + \sum_{i=1}^{n-1} c_i - \sum_{i=1}^{n} \lambda_i \right] = \sum_{i=1}^{n-1} \frac{f(c_i) g(\lambda)}{g'(c_i)(\lambda - c_i)}. \tag{8.3.3}$$

事实上, 因为

$$\frac{g(c_k)}{g'(c_k)(c_k - c_i)} = \begin{cases} 1, & k = i, \\ 0, & k \neq i. \end{cases}$$

于是 (8.3.3) 的两边在 $\lambda = c_1, \cdots, c_{n-1}$ 处取值相同, 但 (8.3.3) 的两边又都是 λ 的 $n-1$ 次多项式, 所以 (8.3.3) 对一切 λ 成立. 于是

$$f(\lambda) = g(\lambda) \left[\lambda + \sum_{i=1}^{n-1} c_i - \sum_{i=1}^{n} l_i \right] + \sum_{i=1}^{n-1} \frac{f(c_i) g(\lambda)}{(c_i)(\lambda - c_i)}. \tag{8.3.4}$$

因为

$$f(c_k) = (-1)^k \prod_{j=1}^{n} |c_k - c_j|,$$

$$g'(c_k) = (-1)^{k-1} \prod_{\substack{j=1 \\ j \neq k}}^{n-1} |c_k - c_j| \neq 0,$$

故 $f(c_k)$ 与 $g'(c_k)$ 异号. 令

$$b_k^2 = \frac{-f(c_k)}{g'(c_k)}, \qquad k = 1, \cdots, n-1, \tag{8.3.5}$$

$$d = \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{n-1} c_j. \tag{8.3.6}$$

于是由 $f(\lambda)$ 的定义及 (8.3.4), (8.3.5) 和 (8.3.6) 得

$$f(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i) = \det(\lambda I - A).$$

这就证明了 $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值.

若 $c_1,\cdots,c_n$ 不是互异的. 我们考虑最简单情况, 即其中只有两个 c_i 是相等的. 不妨设 $c_1=c_2$, 注意在此情况下 $c_1=\lambda_2=c_2$. 剔除 λ_2 和 c_2, 对 $\lambda_1,\lambda_3,\cdots,\lambda_n$ 和 $c_1,c_3,\cdots,c_{n-1}$ 应用前面的方法, 可构造出 $n-1$ 阶实对称阵

$$\tilde{A}=\begin{pmatrix}\tilde{\Lambda}_c & \tilde{b}\\ \tilde{b}' & \tilde{d}\end{pmatrix},$$

它的特征值为 $\lambda_1,\lambda_3,\cdots,\lambda_n$, 其中 $\tilde{\Lambda}_c=\mathrm{diag}(c_1,c_3,\cdots,c_{n-1})$, $\tilde{b}=(b_1,b_3,\cdots,b_{n-1})'$,

$$\tilde{d}=\sum_{\substack{j=1\\ j\neq 2}}^{n}\lambda_j-\sum_{\substack{j=1\\ j\neq 2}}^{n-1}c_j.$$

最后, 定义 $n\times n$ 实对称阵

$$A=\begin{pmatrix}\Lambda & b\\ b' & d\end{pmatrix},$$

其中 $\Lambda=\mathrm{diag}(c_1,c_2,c_3,\cdots,c_{n-1})$, $b'=(b_1,0,b_3,\cdots,b_{n-1})$,

$$d=\tilde{d}=\sum_{j=1}^{n}\lambda_j-\sum_{j=1}^{n-1}c_j.$$

因 $\lambda_2=c_2$, 所以 A 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$. 从上面的讨论, 容易知道对于有多个 c_i 相等时 A 的构造方法. 引理证毕.

引理 8.3.2 设 $x,y\in D$, $x\prec y$. 则存在 $c_1,\cdots,c_{n-1}$, 使得

$$y_1\geqslant c_1\geqslant y_2\geqslant c_2\geqslant\cdots\geqslant c_{n-1}\geqslant y_n,$$

且 $\tilde{x}\prec c$, 这里 $\tilde{x}=(x_1,\cdots,x_{n-1})'$, $c=(c_1,\cdots,c_{n-1})'$.

证明 当 $n=2$ 时, 结论显然成立. 以下假设 $n\geqslant 3$. 记

$$S=\left\{z:\ z\in R^{n-1}, y\geqslant z_1\geqslant y_2\geqslant\cdots\geqslant z_{n-1}\geqslant y_n, \sum_{j=1}^{i}x_j\leqslant\sum_{j=1}^{i}z_j, i=1,\cdots,n-2\right\}.$$

不难验证, S 是一个有界闭凸集. 记

$$m^*=\max_{s}\sum_{i=1}^{n-1}z_i,\quad m_*=\min_{s}\sum_{i=1}^{n-1}z_i.$$

显然 m^* 在 $z=\tilde{y}=(y_1,\cdots,y_{n-1})'$ 达到. 设 m_* 在 $z=\theta=(\theta_1,\cdots,\theta_{n-1})'$ 处达到. 因为

$$m_*=\sum_{i=1}^{n-1}y_i\geqslant\sum_{i=1}^{n-1}x_i,$$

若能证明

$$\sum_{i=1}^{n-1} x_i \geqslant m_*, \tag{8.3.7}$$

则存在 $a \in [0,1]$, 使得

$$am^* + (1-a)m_* = \sum_{i=1}^{n-1} x_i.$$

记 $c = \alpha\tilde{y} + (1-\alpha)\theta$, 由 S 的凸性知, $c \in S$, 且

$$\sum_{i=1}^{n-1} c_i = \sum_{i=1}^{n-1} x_i,$$

因而有 $\tilde{x} \prec c$.

现在证明 (8.3.7). 因 $\theta \in S$, 故有

$$\theta_i \geqslant y_{i+1}, \qquad i = 1, \cdots, n-1, \tag{8.3.8}$$

$$\sum_{j=1}^{i} \theta_j \geqslant \sum_{j=1}^{i} x_j, \qquad i = 1, \cdots, n-1. \tag{8.3.9}$$

因为对 θ_{n-1} 的限制仅仅是 $y_{n-1} \geqslant \theta_{n-1} \geqslant y_n$, 由在 θ 处达到极小值知, $\theta_{n-1} = y_n$. 若在 (8.3.9) 中对所有 $i = 1, \cdots, n-1$, 等号均不成立, 则定义 $l = 0$, 不然则定义

$$l = \max\left\{i\colon \ \sum_{j=1}^{i} \theta_j = \sum_{j=1}^{i} x_j,\ i = 1, \cdots, n-2\right\},$$

当然有 $0 \leqslant l \leqslant n-2$. 另外我们断言

$$\theta_i = y_{i+1},\ 对一切\ i \in \{l+1, \cdots, n-1\}. \tag{8.3.10}$$

事实上, 若 (8.3.10) 不成立, 则存在 $k \in \{l+1, \cdots, n-2\}$. 使 $\theta_k > y_{k+1}$. 取

$$\varepsilon = \min\left\{\sum_{j=1}^{i} \theta_j - \sum_{j=1}^{i} x_j\colon\ i \in \{k+1, \cdots, n-2\}\right\} > 0,$$

则存在 δ, 满足 $0 < \delta < \varepsilon$, 使得

$$y_{k+1} \leqslant \theta_k - \delta \leqslant \theta_k.$$

定义 $\tilde{\theta} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_{n-1})'$, 其中

$$\tilde{\theta}_i = \begin{cases} \theta_i, & i \neq k, \\ \theta_k - \delta, & i = k. \end{cases}$$

易见, $\tilde{\theta} \in S$, 但有

$$\sum_{i=1}^{n-1} \tilde{\theta}_i < \sum_{i=1}^{n-1} \theta_i = m_*,$$

这与 m_* 的定义相矛盾, 这就证明了 (8.3.10). 于是

$$\begin{aligned} m_* = \sum_{i=1}^{n-1} \theta_i &= \sum_{i \leqslant l} \theta_i + \sum_{i=l+1}^{n-1} \theta_i \\ &= \sum_{i \leqslant l} x_i + \sum_{i=l+1}^{n-1} y_{i+1} \qquad (\text{由 (8.3.10) 及 } l \text{ 之定义}) \\ &\leqslant \sum_{i \leqslant l} x_i + \sum_{i=l+2}^{n} x_i \qquad (\text{由 } x < y) \\ &\leqslant \sum_{i \leqslant l} x_i + \sum_{i=l+1}^{n-1} x_i \qquad (\text{由 } x \in D) \\ &= \sum_{i=1}^{n-1} x_i. \end{aligned}$$

(8.3.7) 得证. 引理证毕.

现在我们证明定理 8.3.1 的逆定理.

定理 8.3.2 设 $\alpha, \lambda \in R^n$, 且 $\alpha \prec \lambda$. 则存在实对称矩阵 A, 它以 α 为对角元向量, 以 λ 为特征值向量.

证明 用归纳法来证明. 当 $n = 1$ 时, 定理自然成立. 现设定理对 $n-1$ 成立. 不失一般性, 设 $\alpha = (\alpha_1, \cdots, \alpha_n)' \in D$, $\lambda = (\lambda_1, \cdots, \lambda_n)' \in D$, $\alpha < \lambda$. 由引理 8.3.2 知, 存在 $c = (c_1, \cdots, c_{n-1})'$, 使得

$$\lambda_1 \geqslant c_1 \geqslant \lambda_2 \geqslant c_2 \geqslant \cdots \geqslant c_{n-1} \geqslant \lambda_n,$$

且 $\tilde{\alpha} < c$, 这里 $\tilde{\alpha} = (\alpha_1, \cdots, \alpha_{n-1})$. 应用归纳假设, 存在 $n-1$ 阶实对称阵 A_1, 使得 A_1 以 $\tilde{a}$ 为对角元向量, 以 c 为特征值向量. 同时由定理 1.3.1 知, 存在 $n-1$ 阶正交阵 Q, 使 $A_1 = QD_cQ'$, 其中 $D_c = \text{diag}(c_1, \cdots, c_{n-1})$.

应用引理 8.3.1, 我们有 n 阶实对称阵

$$\tilde{A} = \begin{pmatrix} D_c & b \\ b' & d \end{pmatrix}$$

以 λ 为特征值向量, 这里

$$d = \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n-1} c_i = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n-1} \alpha_i = \alpha_n.$$

最后定义 A 为

$$A=\begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} D_c & b \\ b' & d \end{pmatrix}\begin{pmatrix} Q' & 0 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} A_1 & Qb \\ b'Q' & d \end{pmatrix}.$$

显然 A 与 $\tilde{A}$ 相似, 所以 A 的特征值向量也为 λ. 因 A_1 以 $\tilde{a}=(a_1,\cdots,a_{n-1})'$ 为对角元向量, 又 $d=a_n$, 所以 A 的对角元向量为 a. 定理证毕.

下面的定理建立了 Hermite 阵的特征值向量与主子阵的特征值向量之间的受控关系.

定理 8.3.3　设 A 为 $n\times n$ Hermite 阵, 分块为

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

其中 A_{11} 和 A_{22} 分别为 $r\times r$ 和 $(n-r)\times(n-r)$ 方阵. 则

$$\begin{pmatrix} \lambda(A_{11}) \\ \lambda(A_{22}) \end{pmatrix}\prec\lambda(A),$$

其中 $\lambda(A)$ 表示 A 的特征值组成的 $n\times 1$ 向量.

证明　设 $\alpha_1,\cdots,\alpha_r$ 和 $\beta_1,\cdots,\beta_{n-r}$ 分别为 A_{11} 和 A_{22} 的特征值. 记

$$\Lambda_\alpha=\mathrm{diag}(\alpha_1,\cdots,\alpha_r),$$

$$\Lambda_\beta=\mathrm{diag}(\beta_1,\cdots,\beta_{n-r}).$$

由定理 1.4.1 知, 存在酉阵 U_1 和 U_2, 使得

$$U_1^*A_{11}U_1=\Lambda_\alpha,\quad U_2^*A_{22}U_2=\Lambda_\beta.$$

记 $U=\mathrm{diag}(U_1,U_2)$, 则

$$U^*AU=\begin{pmatrix} \Lambda_\alpha & M \\ M^* & \Lambda_\beta \end{pmatrix},$$

其中 $M=U_1^*A_{12}U_2$, 注意到 $\alpha_1,\cdots,\alpha_r$, $\beta_1,\cdots,\beta_{n-r}$ 为 U^*AU 的对角元. 由定理 8.3.1 得

$$\begin{pmatrix} \lambda(A_{11}) \\ \lambda(A_{22}) \end{pmatrix}=(\alpha_1,\cdots,\alpha_r,\ \beta_1,\cdots,\beta_{n-1})'<\lambda(U^*AU)$$

$$=\lambda(A).$$

证毕.

下一个定理是定理 8.3.3 的进一步推广, 为了证明它, 我们需要如下引理.

引理 8.3.3 设 A 为 $n \times n$ Hermite 阵, $\alpha_1 \geqslant \cdots \geqslant \alpha_n$ 为其特征值. 记

$$g_k(A) = \sum_{i=k}^{n} \alpha_i,$$

$$h_k(A) = \sum_{i=1}^{k} \alpha_i,$$

则 $g_k(A)$ 为 A 的凹函数, $h_k(A)$ 为 A 的凸函数.

证明 设 $\beta_1 \geqslant \cdots \geqslant \beta_n$, $\gamma_1 \geqslant \cdots \geqslant \gamma_n$ 分别为 B 和 $\theta A + (1-\theta)B$ 的特征值, 这里 B 也为 $n \times n$ Hermite 阵, $\theta \in [0,1]$. 记

$$\tilde{A} = I + \varepsilon A,$$

$$\tilde{B} = I + \varepsilon B.$$

则当 $\varepsilon > 0$ 充分小时, $\tilde{A} > 0$, $\tilde{B} > 0$. 由推论 3.9.1, 有

$$\prod_{i=k}^{n} \lambda_i(\theta\tilde{A} + (1-\theta)\tilde{B}) \geqslant \prod_{i=k}^{n} (1+\varepsilon\alpha_i)^{\theta}(1+\varepsilon\beta_i)^{1-\theta}, \quad 0 \leqslant \theta \leqslant 1.$$

因为 $\theta\tilde{A} + (1-\theta)\tilde{B} = I + \varepsilon(\theta A + (1-\theta)B)$, 上式即为

$$\prod_{i=k}^{n} (1+\varepsilon\gamma_i) \geqslant \prod_{i=k}^{n} (1+\varepsilon\alpha_i)^{\theta}(1+\varepsilon\beta_i)^{1-\theta}.$$

将上式两边展成 ε 的多项式, 并将 ε 的高于一次的项合并, 得到

$$1 + \varepsilon\sum_{i=k}^{n} \gamma_i + O(\varepsilon^2) \geqslant 1 + \theta\varepsilon\sum_{i=k}^{n} \alpha_i + (1-\theta)\varepsilon\sum_{i=k}^{n} \beta_i + O(\varepsilon^2),$$

等价地

$$\sum_{i=k}^{n} \gamma_i + O(\varepsilon) \geqslant \theta\sum_{i=k}^{n} \alpha_i + (1-\theta)\sum_{i=k}^{n} \beta_i + O(\varepsilon).$$

命 $\varepsilon \to 0$, 便有

$$g_k(\theta A + (1-\theta)B) \geqslant \theta g_k(A) + (1-\theta)g_k(B).$$

这就证明了 $g_k(A)$ 的凹性.

再利用 $h_k(A) = -g_{n-k+1}(-A)$, 便可推出 $h_k(A)$ 的凸性. 引理证毕.

注 1 特别当$k=1$, $h_1(A)=\lambda_1(A)$是A的凸函数. 即设A和B为两个n阶 Hermite 阵, 则对 $\theta(0\leqslant\theta\leqslant1)$,

$$\lambda_1(\theta A+(1-\theta)B)\leqslant\theta\lambda_1(A)+(1-\theta)\lambda_1(B).$$

定理 8.3.4 设 A 如定理 8.3.3, 记

$$A_\theta=\begin{pmatrix}A_{11} & \theta A_{12}\\ \theta A_{21} & A_{22}\end{pmatrix},\qquad \theta\in[0,1],$$

则 $\lambda(A_{\theta_1})\prec\lambda(A_{\theta_2})$, $0\leqslant\theta_1\leqslant\theta_2\leqslant1$.

证明 因为我们可把 A_{θ_2} 当作 A 来讨论, 故只需对 $\theta_2=1$ 证明本定理, 即证明 $\lambda(A_\theta)\prec\lambda(A)$, 其中 $0\leqslant\theta\leqslant1$.

因 $A_\theta=\theta A+(1-\theta)A_0$, $\theta\in[0,1]$, 利用引理 8.3.3 中 $h_k(A)$ 的凸性, 得

$$\sum_{i=1}^{k}\lambda_i(A_\theta)\leqslant\theta\sum_{i=1}^{k}\lambda_i(A)+(1-\theta)\sum_{i=1}^{k}\lambda_i(A_0).\tag{8.3.11}$$

但从定理 8.3.3, 我们有 $\lambda(A_0)\prec\lambda(A)$, 因而

$$\sum_{i=1}^{k}\lambda_i(A_0)\leqslant\sum_{i=1}^{k}\lambda_i(A),$$

代入 (8.3.11) 得

$$\sum_{i=1}^{k}\lambda_i(A_\theta)\leqslant\sum_{i=1}^{k}\lambda_i(A),$$

即我们证明了 $\lambda(A_\theta)\prec\lambda(A)$. 证毕.

注 2 当 $\theta_1=0$, $\theta_2=1$, 此定理变为定理 8.3.3.

在这一节的最后部分, 我们讨论涉及两个 Hermite 之和的特征值向量的几个受控关系.

定理 8.3.5 设 A 和 B 为两个 $n\times n$ 半正定 Hermite 阵, 则

$$\begin{pmatrix}\lambda(A)\\ \lambda(B)\end{pmatrix}\prec\begin{pmatrix}\lambda(A+B)\\ 0\end{pmatrix}.$$

证明 因为 A 和 B 皆为半正定 Hermite 阵, 由定理 1.4.2 知, 存在 n 阶方阵 P 与 Q, 使得 $A=PP^*$, $B=QQ^*$. 记 $C=(P\vdots Q)$, 则 $A+B=CC^*$, 且

$$C^*C=\begin{pmatrix}P^*P & P^*Q\\ Q^*P & Q^*Q\end{pmatrix}\triangleq D.$$

应用定理 8.3.3, 得

$$\begin{pmatrix} \lambda(P^*P) \\ \lambda(Q^*Q) \end{pmatrix} \prec \lambda(D).$$

但 $\lambda(P^*P)=\lambda(A)$, $\lambda(Q^*Q)=\lambda(B)$,

$$\lambda(D)=\lambda(C^*C)=\begin{pmatrix} \lambda(A+B) \\ 0 \end{pmatrix}.$$

证毕.

定理 8.3.6 设 A 和 B 为两个 n 阶 Hermite 阵. 记 $\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$, $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$ 为它们的特征值. 则

$$\lambda(A+B)\prec\lambda(A)+\lambda(B). \tag{8.3.12}$$

证明 应用定理 6.6.1, 有

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i(A+B) &= \max_{X^*X=I_k}\text{tr}X^*(A+B)X \\ &= \max_{X^*X=I_k}[\text{tr}X^*AX+\text{tr}X^*BX] \\ &\leqslant \max_{X^*X=I_k}\text{tr}X^*AX+\max_{X^*X=I_k}\text{tr}X^*BX \\ &= \sum_{i=1}^{k}[\lambda_i(A)+\lambda_i(B)].\end{aligned}$$

证毕.

在这个定理中, 若 $B=bb^*$, 其中 b 为 $n\times1$ 复向量, 则有 $\lambda_1(B)=b^*b$, $\lambda_i(B)=0$, $i\geqslant2$. 于是我们得到如下推论.

推论 8.3.5 设 A 为 $n\times n$ Hermite 阵, $C=A+bb^*$, 其中 b 为 $n\times1$ 复向量. 则

$$\lambda(C)\prec(\lambda_1(A)+b^*b,\lambda_2(A),\cdots,\lambda_n(A))'. \tag{8.3.13}$$

定理 8.3.7(Fan-Hoffman 不等式) 设 A 为 $n\times n$ 半正定阵, 则对任意酉矩阵 U 有

$$\lambda(A-I)\prec_w\lambda(A-U)\prec\lambda(A+I). \tag{8.3.14}$$

证明参见 Fan 和 Hoffman(1955) 或 Marshall 和 Olkin(1979) p. 266.

Wang Boying, Xi Boyan 和 Zhang Fuzhen (1999) 指出对一般形式

$$\lambda(A-B)\prec_w\lambda(A-BU)\prec\lambda(A+B) \tag{8.3.15}$$

未必成立, 这里 B 为 $n\times n$ 半正定阵. 如令

$$A=\begin{pmatrix}5&2\\2&4\end{pmatrix},\quad B=\begin{pmatrix}1&2\\2&4\end{pmatrix},\quad U=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

则 $\lambda_1(A-B)=4>\lambda_1(A-BU)=3.6226$. 但去掉其中间项后, (8.3.15) 成立, 即如下定理.

定理 8.3.8　设 A, B 为 $n\times n$ 半正定阵, 则有

$$\lambda(A-B)\prec\lambda(A+B). \tag{8.3.16}$$

定理的证明可由下节的推论 8.7.1 和注 1 直接得出.

§8.4　一般复方阵

在上节, 我们对 Hermite 阵, 建立了对角元向量和特征值向量之间的受控关系. 但对于一般复方阵, 因为它的对角元和特征值一般都是复数, 所以在研究受控关系时转而考虑它们的实部或模组成的向量. 遗憾的是, 对定理 8.3.1 的结果, 把 A 换成一般复方阵, α, λ 换成它们的实部组成的向量, 相应的受控关系并不成立. 一个简单例子是

$$A=\frac{1}{5}\begin{pmatrix}7+3i&1+4i\\-4+4i&3-3i\end{pmatrix}=\frac{1}{5}\begin{pmatrix}2&1\\1&-2\end{pmatrix}\begin{pmatrix}1+i&0\\0&1-i\end{pmatrix}\begin{pmatrix}2&1\\1&-2\end{pmatrix}.$$

A 的两个特征值为 $1+i$ 和 $1-i$, 故 $(\mathrm{Re}\lambda_1,\mathrm{Re}\lambda_2)=(1,1)$. 而 $(\mathrm{Re}a_{11},\mathrm{Re}a_{22})=\left(\frac{7}{5},\frac{3}{5}\right)$, 有 $(\mathrm{Re}\lambda_1,\mathrm{Re}\lambda_2)'\prec(\mathrm{Re}a_{11},\mathrm{Re}a_{22})'$. 如果用奇异值代替特征值, 我们可以有下列的受控关系.

记 $\sigma_1(A)\geqslant\sigma_2(A)\geqslant\cdots\geqslant\sigma_n(A)$ 为 A 的奇异值.

定理 8.4.1　设 $A=(\alpha_{ij})$ 为 $n\times n$ 复方阵. 记

$$\mathrm{Re}(\alpha)=(\mathrm{Re}\alpha_{11},\cdots,\mathrm{Re}\alpha_{nn})',$$

$$\sigma(A)=(\sigma_1(A),\cdots,\sigma_n(A))',$$

$$|\alpha|=(|\alpha_{11}|,\cdots,|\alpha_{nn}|)'.$$

则

$$\mathrm{Re}(a)\prec_w|\alpha|\prec_w\sigma(A). \tag{8.4.1}$$

证明　对 A 作奇异值分解 (见定理 1.5.4), 得

$$A=U\varDelta V^*,$$

这里 U 和 V 为酉阵, $\Delta = \mathrm{diag}(\sigma_1(A), \cdots, \sigma_n(A))$, 则

$$a_{ii} = \sum_j u_{ij} v_{ij}^* \sigma_j(A).$$

于是

$$|\alpha_{ii}| \leqslant \sum_j |u_{ij} v_{ij}^*| \sigma_j(A) = \sum_j p_{ij} \sigma_j(A), \tag{8.4.2}$$

其中 $p_{ij} = |u_{ij} v_{ij}^*|$. 定义 $P = (p_{ij})$, 则 P 为次双随机阵. 记

$$c_i = \frac{|\alpha_{ii}|}{\sum\limits_j p_{ij} \sigma_j(A)}, \qquad i = 1, \cdots, n,$$

则 $0 \leqslant c_i \leqslant 1$, $i = 1, \cdots, n$. 再命 $\Lambda_c = \mathrm{diag}(c_1, \cdots, c_n)$. 则由 (8.4.2) 得 $|\alpha| = P\Lambda_c \sigma(A)$. 显然 $P\Lambda_c$ 为次双随机阵. 由定理 8.1.6 得 $|\alpha| \prec_w \sigma(A)$. 这就证明了 (8.4.1) 的第二个关系. 第一个受控关系是显然的, 定理证毕.

根据奇异值的定义, Hermite 阵的奇异值等于其特征值的绝对值. 对于一般复方阵, 当然这个结论并不成立, 然而我们有如下结果.

定理 8.4.2 设 A 为 $n \times n$ 复方阵, 其特征值按它们的模做递减排列: $|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n(A)|$. 则

$$\prod_{i=1}^k |\lambda_i(A)| \leqslant \prod_{i=1}^k \sigma_i(A), \quad k = 1, \cdots, n-1, \tag{8.4.3}$$

$$\prod_{i=1}^n |\lambda_i(A)| = \prod_{i=1}^n \sigma_i(A). \tag{8.4.4}$$

当 $|\lambda_n(A)| > 0$ 时, 上两式等价于

$$(\log |\lambda_1(A)|, \cdots, \log |\lambda_n(A)|)' \prec (\log \sigma_1(A), \cdots, \log \sigma_n(A))'. \tag{8.4.5}$$

证明 设 x 为 A 的对应于特征值 $\lambda_1(A)$ 的单位特征向量, 于是 $Ax = \lambda_1(A)x$. 应用定理 4.1.1, 有

$$|\lambda_1(A)|^2 = \lambda_1(A)\overline{\lambda_1(A)} = x^* A^* A x \leqslant \lambda_1(A^* A) = \sigma_1^2(A),$$

即

$$|\lambda_1(A)| \leqslant \sigma_1(A). \tag{8.4.6}$$

记 $A^{(k)}$ 表示 A 的第 k 次复合 (关于 "复合" 的定义及性质见本章末的附录). 对 $A^{(k)}$ 应用 (8.4.6) 得

$$|\lambda_1(A^{(k)})| \leqslant \sigma_1(A^{(k)}). \tag{8.4.7}$$

根据复合阵的性质 (见定理 8.9.2), 有

$$|\lambda_1(A^{(k)})| = \left|\prod_{i=1}^{k} \lambda_i(A)\right|, \qquad k = 1, \cdots, n,$$

$$\sigma_1(A^{(k)}) = \prod_{i=1}^{k} \sigma_i(A), \qquad k = 1, \cdots, n.$$

结合 (8.4.7), 定理得证.

推论 8.4.1　对任意 n 阶复方阵 A, 有

(1) $(|\lambda_1(A)|, \cdots, |\lambda_n(A)|)' \prec_w (\sigma_1(A), \cdots, \sigma_n(A))'$;

(2) $(|\lambda_1(A)|^2, \cdots, |\lambda_n(A)|^2)' \prec_w (\sigma_1^2(A), \cdots, \sigma_n^2(A))'$.

证明　(1) 对 (8.4.5) 应用定理 8.2.13, 取 $g(z) = e^z$, 便得欲证; 将 (1) 与例 8.2.8(1) 相结合便得到 (2). 证毕.

推论 8.4.2(Schur)　对任意 n 阶复方阵 $A = (\alpha_{ij})$, 有

$$\sum_{i=1}^{n} |\lambda_i(A)|^2 \leqslant \sum_{ij} |\alpha_{ij}|^2 = \mathrm{tr}AA^* = \sum_{i=1}^{n} \lambda_i(AA^*).$$

这是推论 8.4.1 的直接结果. 该事实我们在 §4.7 用另外的方法证明过 (见推论 4.7.1).

我们用 $\sigma(A)$ 表示方阵 A 的奇异值组成的向量. 约定 $\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$ 和 $\sigma(A) + \sigma(B) = (\sigma_1(A) + \sigma_1(B), \cdots, \sigma_n(A) + \sigma_n(B))'$.

下面的定理建立了两个复方阵的奇异值向量与它们之和的奇异值向量之间的受控关系.

定理 8.4.3　设 A 和 B 为两个复方阵, 则

$$\sigma(A + B) \prec_w \sigma(A) + \sigma(B).$$

证明　对 Hermite 阵

$$\tilde{A} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \text{ 和 } \tilde{B} = \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}$$

应用定理 8.3.6, 得

$$\lambda(\tilde{A} + \tilde{B}) \prec \lambda(\tilde{A}) + \lambda(\tilde{B}). \tag{8.4.8}$$

注意到 $\tilde{A}$ 的非零特征值是 A 的非零奇异值及它们的相反数, 所以从 (8.4.8) 即得 $\sigma(A + B) \prec_w \sigma(A) + \sigma(B)$. 证毕.

推论 8.4.3　对任意复方阵 A 和 B, 有

$$\sigma_1(A + B) \leqslant \sigma_1(A) + \sigma_1(B).$$

§8.5 复方阵的 Hermite 部分

设 A 为一复方阵, 方阵

$$H=\frac{1}{2}(A+A^*)$$

称为 A 的 Hermite 部分. 当然 H 的特征值都是实数. 本节给出 H, A 及 AA^* 特征值之间的受控关系.

定理 8.5.1 设 A 为 $n\times n$ 复方阵, 则

$$(\mathrm{Re}\lambda_1(A),\cdots,\mathrm{Re}\lambda_n(A))'\prec(\lambda_1(H),\cdots,\lambda_n(H))'.$$

证明 根据 Schur 三角化定理 (见定理 1.5.7), $A=ULU^*$, 这里 U 为 $n\times n$ 酉阵, L 为下三角阵, 且 $l_{ii}=\lambda_i(A), i=1,\cdots,n$. 假设 H 的特征值和 A 的特征值的实部都按递减排列

$$\lambda_1(H)\geqslant\cdots\geqslant\lambda_n(H),$$

$$\mathrm{Re}\lambda_1(A)\geqslant\cdots\geqslant\mathrm{Re}\lambda_n(A),$$

依定理 6.6.1, 有

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i(H)&=\max_{X^*X=I_k}\mathrm{tr}X^*HX\\&=\max_{X^*X=I_k}\mathrm{tr}X^*U\left(\frac{L+L^*}{2}\right)U^*X\\&\geqslant\mathrm{tr}(I_k 0)\left(\frac{L+L^*}{2}\right)\begin{pmatrix}I_k\\0\end{pmatrix}\\&=\sum_{i=1}^{k}\frac{l_{ii}+\overline{l_{ii}}}{2}=\sum_{i=1}^{k}\mathrm{Re}\lambda_i(A).\end{aligned}$$

明所欲证.

推论 8.5.1 对任意复方阵 A, 记 $\lambda(A)$ 为其任一特征值, 则

$$\mathrm{Re}\lambda(A)\leqslant\lambda_1\left(\frac{A+A^*}{2}\right).$$

定理 8.5.1 的逆命题也成立, 这就是

定理 8.5.2 设 $z_1,\cdots,z_n$ 为复数, $\alpha_1,\cdots,\alpha_n$ 为实数, 满足

$$(\mathrm{Re}z_1,\cdots,\mathrm{Re}z_n)'\prec(\alpha_1,\cdots,\alpha_n)',$$

则存在 n 阶复方阵 A, 以 $z_1,\cdots,z_n$ 为特征值, 同时它的 Hermite 部分以 $\alpha_1,\cdots,\alpha_n$ 为特征值.

证明　根据定理 8.3.2, 知存在实对称阵 B, 以 $\mathrm{Re}z_1,\cdots,\mathrm{Re}z_n$ 为对角元, 以 $\alpha_1,\cdots,\alpha_n$ 为特征值. 由定理 1.3.1, 又存在正交阵 Q, 使得 $Q'BQ=\Lambda_\alpha$, 其中 $\Lambda_\alpha=\mathrm{diag}(\alpha_1,\cdots,\alpha_n)$, 设 $L=(l_{ij})$ 为下三角阵, 这里

$$l_{ij}=\begin{cases} z_i, & i=j, \\ 2b_{ij}, & i>j, \\ 0, & i<j. \end{cases} \quad \text{其中 } i,j=1,\cdots,n.$$

显然, $z_i=\lambda_i(L)=\lambda_i(Q'LQ)$. 定义 $A=Q'LQ$, 则

$$\frac{A+A^*}{2}=Q'\left(\frac{L+L^*}{2}\right)Q=Q'BQ=\Lambda_\alpha.$$

于是这样定义 A 就是所需的方阵. 证毕.

§8.6　矩阵乘积

设 A 为任一 $n\times n$ 复方阵, 记 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$ 为其奇异值.

定理 8.6.1　设 A 和 B 为 $n\times n$ 复方阵, 则

$$\prod_{i=1}^{k}\sigma_i(AB)\leqslant\prod_{i=1}^{k}\sigma_i(A)\sigma_i(B),\quad k=1,\cdots,n-1,$$

$$\prod_{i=1}^{n}\sigma_i(AB)=\prod_{i=1}^{n}\sigma_i(A)\sigma_i(B).$$

当 $\sigma_n(AB)>0$ 时, 等价地

$$(\log\sigma_1(AB),\cdots,\log\sigma_n(AB))'\prec(\log\sigma_1(A)+\log\sigma_1(B),\cdots,\log\sigma_n(A)+\log\sigma_n(B))'.$$

因为 $\sigma_i(A)=\lambda_i^{1/2}(AA^*)$, $\sigma_i(AB)=\lambda_i^{1/2}(ABB^*A^*)=\lambda_i^*(A^*ABB^*)$, $i=1,\cdots,n$, 于是上面定理的等价形式为

定理 8.6.2　设 A 和 B 为 $n\times n$ 半正定 Hermite 阵, 则

$$\prod_{i=1}^{k}\lambda_i(AB)\leqslant\prod_{i=1}^{k}\lambda_i(A)\lambda_i(B),\quad i=1,\cdots,n-1, \tag{8.6.1}$$

$$\prod_{i=1}^{n}\lambda_i(AB)=\prod_{i=1}^{n}\lambda_i(A)\lambda_i(B), \tag{8.6.2}$$

当 $\lambda_n(AB) > 0$ 时, 等价地

$$(\log\lambda_1(AB), \cdots, \log\lambda_n(AB))' \prec (\log\lambda_1(A) + \log\lambda_1(B), \cdots, \log\lambda_n(A) + \log\lambda_n(B))'. \tag{8.6.3}$$

证明 由推论 4.6.3, 得

$$\lambda_1(AB) \leqslant \lambda_1(A)\lambda_1(B). \tag{8.6.4}$$

考虑 A 和 B 的第 k 次复合 $A^{(k)}$ 和 $B^{(k)}$, 因为

$$(AB)^{(k)} = A^{(k)}B^{(k)},$$

(见定理 8.9.1), 利用 (8.6.4), 我们有

$$\lambda_1((AB)^{(k)}) \leqslant \lambda_1(A^{(k)})\lambda_1(B^{(k)}). \tag{8.6.5}$$

注意到

$$\lambda_1((A)^{(k)}) = \prod_{i=1}^{k}\lambda_i(A),$$

$$\lambda_1(B^{(k)}) = \prod_{i=1}^{k}\lambda_i(B),$$

$$\lambda_1((AB)^{(k)}) = \prod_{i=1}^{k}\lambda_i(AB),$$

代入 (8.6.5) 即得 (8.6.1). (8.6.2) 是显然的. 证毕.

推论 8.6.1 在定理假设下, 有

$$(\lambda_1(AB), \cdots, \lambda_n(AB))' \prec (\lambda_1(A)\lambda_1(B), \cdots, \lambda_n(A)\lambda_n(B))'.$$

结论由 (8.6.3) 和定理 (8.2.13) 直接推出.

最后, 我们证明两个关于 AB 的若干个特征值之和的一个不等式.

定理 8.6.3 设 A 和 B 为两个 $n \times n$ 半正定 Hermite 阵, 则

$$\sum_{i=1}^{k}\lambda_i(A)\lambda_{n-i+1}(B) \leqslant \sum_{i=1}^{k}\lambda_i(AB) \leqslant \sum_{i=1}^{k}\lambda_i(A)\lambda_i(B).$$

证明 右边的不等式由推论 8.6.1 直接推得. 下证左边不等式.

不失一般性, 可设 A 为对角阵, 设 $A = \Lambda_n = \text{diag}(\alpha_1, \cdots, \alpha_n)$, 其中 $\alpha_i = \lambda_i(A) \geqslant 0$. 再记 $\Lambda_k = \text{diag}\ (\alpha_1, \cdots, \alpha_k, 0, \cdots, 0)$. 记 U 为 $n \times k$ 矩阵, 满足

$U^*U = I_k$, 应用定理 6.6.1, 得

$$\begin{aligned}\sum_{i=1}^k \lambda_i(AB) &= \sum_{i=1}^k \lambda_i(\Lambda_n^{1/2} B \Lambda_n^{1/2}) \\ &= \max_U \mathrm{tr} U \Lambda_n^{1/2} B \Lambda_n^{1/2} U^* \\ &\geqslant \mathrm{tr}(I_k \vdots 0) \Lambda_n^{1/2} B \Lambda_n^{1/2} \begin{pmatrix} I_k \\ 0 \end{pmatrix} \\ &= \mathrm{tr} \Lambda_k^{1/2} B \Lambda_k^{1/2} = \mathrm{tr} \Lambda_k B \\ &\geqslant \sum_{i=1}^k \lambda_i(A) \lambda_{n-i+1}(B),\end{aligned}$$

最后一个不等号是由于 Neumann 不等式 (定理 4.6.1). 定理证毕.

下面的定理是定理 8.6.3 的进一步改进, 它是由 Wang Boying (王伯英) 和 Fuzhen Zhang(张福基)(1992) 证明的.

定理 8.6.4 设 A 和 B 为 $n \times n$ 半正定 Hermite 阵. 则

$$\sum_{t=1}^k \lambda_{i_t}(AB) \geqslant \sum_{t=1}^k \lambda_{i_t}(A) \lambda_{n-t+1}(B), \quad 1 \leqslant i_1 < \cdots < i_k \leqslant n.$$

证明 设 $S_1 \subset \cdots \subset S_k$ 为 C^n 的子空间, 它们的维数分别为 $i_1, \cdots, i_k$. 取 $u_i \in S_i$, $i = 1, \cdots, k$, 它们为标准正交化向量. 再记 $U = (u_1, \cdots, u_k)$, 则 $U^*U = I_k$. 应用引理 4.4.1, 得

$$\sum_{t=1}^k \lambda_{i_t}(B) = \max_{\substack{s_i \\ i=1,\cdots,k}} \min_{U^*U=I_k} \mathrm{tr} U^* B U,$$

因而

$$\begin{aligned}\sum_{t=1}^k \lambda_{i_t}(AB) &= \sum_{t=1}^k \lambda_{i_t}(A^{1/2} B A^{1/2}) \\ &= \max_{\substack{s_i \\ i=1,\cdots,k}} \min_{U^*U=I_k} \mathrm{tr} U^* A^{1/2} B A^{1/2} U. \end{aligned} \tag{8.6.6}$$

记 $\varphi_1, \cdots, \varphi_n$ 为 A 的对应于 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ 的标准正交化特征向量, $S_t = \mathscr{M}(\varphi_1, \cdots, \varphi_{i_t})$, 显然 $S_1 \subset \cdots \subset S_k$. 利用上一定理, 我们有

$$\begin{aligned}\mathrm{tr} U^* A^{1/2} B A^{1/2} U &= \mathrm{tr} B A^{1/2} U U^* A^{1/2} \\ &\geqslant \sum_{i=1}^n \lambda_{n-i+1}(B) \lambda_i(U^* A U)\end{aligned}$$

$$\geqslant \sum_{i=1}^{k} \lambda_{n-i+1}(B)\lambda_i(U^*AU),$$

结合 (8.6.6) 得

$$\sum_{t=1}^{k} \lambda_{i_t}(AB) \geqslant \min_{U^*U=I_k} \sum_{i=1}^{k} \lambda_{n-i+1}(B)\lambda_i(U^*AU). \tag{8.6.7}$$

选 $g_{t+1},\cdots,g_{i_t} \in S_t$, 使得 $u_1,\cdots,u_t$, $g_{i+1},\cdots,g_{i_t}$ 为 S_t 的一组标准正交基. 记 $U_1=(u_1,\cdots,u_t)$, $G=(g_{t+1},\cdots,g_{i_t})$, $\boldsymbol{\Phi}_{i_t}=(\varphi_1,\cdots,\varphi_{i_t})$, 则存在 $i_t \times i_t$ 的酉阵 W, 使得

$$(U_1 \vdots G) = \boldsymbol{\Phi}_{i_t} W. \tag{8.6.8}$$

再分块 $U=(U_1 \vdots U_2)$. 注意到 $U_1^*AU_1$ 是 U^*AU 的主子式, 故应用 Sturm 定理 (定理 4.5.1), 得

$$\lambda_t(U^*AU) \geqslant \lambda_t(U_1^*AU_1), \quad t=1,\cdots,k.$$

另一方面, 因为

$$\begin{pmatrix} U_1^* \\ G^* \end{pmatrix} A(U_1 \vdots G) = \begin{pmatrix} U_1^*AU_1 & U_1^*AG \\ G^*AU_1 & G^*AG \end{pmatrix},$$

同样, 由 Sturm 定理, 并利用 (8.6.8), 得

$$\begin{aligned}\lambda_t(U_1^*AU_1) &\geqslant \lambda_{i_t}\left(\begin{pmatrix} U_1^* \\ G^* \end{pmatrix} A(U_1 \vdots G)\right) \\ &= \lambda_{i_t}(W^*\boldsymbol{\Phi}_{i_t}^* \, A\boldsymbol{\Phi}_{i_t}W) \\ &= \lambda_{i_t}(W^*\mathrm{diag}(\lambda_1(A),\cdots,\lambda_{i_t}(A))W) \\ &= \lambda_{i_t}(A).\end{aligned}$$

代入 (8.6.7), 定理得证.

推论 8.6.2 设 A 和 B 为 $n\times n$ 复方阵, 则

$$\sum_{t=1}^{k} \sigma_{i_t}^2(AB) \geqslant \sum_{t=1}^{k} \sigma_{i_t}^2(A)\sigma_{n-t+1}^2(B).$$

§8.7 Log- 弱受控不等式

在研究复矩阵时, 我们涉及到许多关于特征值和奇异值的积形式的不等式, 如本章 §8.4 和 §8.6 所见. 可以看出采用定义 8.1.1 来描述这种受控关系, 记号比较复

杂, 本节我们将介绍另一种受控概念: Log- 弱受控以及一些重要的 Log- 弱受控不等式.

定义 8.7.1　设对任意两个向量 $x = (x_1, \cdots, x_n)'$, $y = (y_1, \cdots, y_n)'$, 其中 $x_i \geqslant 0$, $y_i \geqslant 0$, 且 $x_1, \geqslant \cdots \geqslant x_n$, $y_1 \geqslant \cdots \geqslant y_n$, 若

$$\prod_{i=1}^{k} x_i \leqslant \prod_{i=1}^{k} y_i, \quad k = 1, \cdots, n, \tag{8.7.1}$$

则称 x Log- 弱受控于 y, 记作 $\log x \prec_w \log y$. 若当 $k = n$ 时, 上式等号成立, 即

$$\prod_{i=1}^{n} x_i \leqslant \prod_{i=1}^{n} y_i, \tag{8.7.2}$$

则称 x Log- 受控于 y, 记作 $\log x \prec \log y$.

对任意复矩阵 A, 设 $\lambda_1(A), \cdots, \lambda_n(A)$ 和 $\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$ 分别为 A 的特征值和奇异值, 其特征值按它们的模作递减排列 $|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n|$, 记

$$\lambda(A) = (\lambda_1(A), \cdots, \lambda_n(A))',$$

$$|\lambda(A)| = (|\lambda_1(A)|, \cdots, |\lambda_n(A)|)',$$

$$\sigma(A) = (\sigma_1(A), \cdots, \sigma_n(A))',$$

我们可得到下列不等式.

定理 8.7.1　设 A, B 和 C 分别为 $n \times n$, $m \times n$ 和 $m \times m$ 复矩阵, $\mathrm{r}(B) = r$, 且

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geqslant 0,$$

则

$$\log \sigma(B) \prec_w \log \mu, \tag{8.7.3}$$

这里 $\mu = (\mu_1, \cdots, \mu_n)'$, 其中

$$\mu_i = \begin{cases} \max\{\lambda_i(A), \lambda_i(C)\}, & 若\ i \leqslant r, \\ 0, & 其他. \end{cases}$$

证明　当 $B = 0$ 时, 显然成立, 下面我们假设 $B \neq 0$. 对 B 进行奇异值分解得

$$B = UDV^*,$$

其中 $D = \mathrm{diag}(\sigma(B), \cdots, \sigma(B))$, U 和 V 分别为 $m \times r$ 和 $n \times r$ 的矩阵, 且满足 $U^*U = V^*V = I_r$. 因此

$$\begin{pmatrix} U^* & 0 \\ 0 & V^* \end{pmatrix}\begin{pmatrix} A & B \\ B^* & C \end{pmatrix}\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} U^*AU & D \\ D & V^*CV \end{pmatrix} \geqslant 0,$$

取每个子块的 k 阶顺序主子阵, $0 \leqslant k \leqslant r$, 我们有

$$\begin{pmatrix} (U^*AU)_k & (D)_k \\ (D)_k & (CV^*CV)_k \end{pmatrix} \geqslant 0,$$

对每个子块求行列式得到

$$\left|(D)_k\right| \leqslant \left|(U^*AU)_k\right|\left|(CV^*CV)_k\right|, \tag{8.7.4}$$

即对任意 $k(0 \leqslant k \leqslant r)$, 有

$$\prod_{i=1}^{k} \sigma_i^2(B) \leqslant \prod_{i=1}^{k} \lambda_i\Big((U^*AU)_k\Big)\lambda_i\Big((V^*CV)_k\Big).$$

由 Poincaré分离定理 (定理 4.6.1), 立得

$$\prod_{i=1}^{k} \sigma_i^2(B) \leqslant \prod_{i=1}^{k} \lambda_i(A)\lambda_i(C) \leqslant \prod_{i=1}^{k} \mu_i^2. \tag{8.7.5}$$

定理证毕.

推论 8.7.1 设 A 和 B 为 n 阶半正定阵, z 为任意复数, 则

$$\log\sigma(A - |z|B) \prec_w \log\sigma(A + zB) \prec_w \log\sigma(A + |z|B). \tag{8.7.6}$$

证明 注意到

$$\begin{pmatrix} A + |z|B & A + zB \\ A + z^*B & A + |z|B \end{pmatrix} \geqslant 0,$$

由定理 8.7.1 立得上式的第二个不等式. 关于第一个不等式的证明参见 Zhan(2000). 推论证毕.

在 (8.7.6) 中取 $z = 1$, 便得不等式

$$\log\lambda(A - B) \prec_w \log\lambda(A + B). \tag{8.7.7}$$

推论 8.7.2(Weyl 不等式) 对任意的 n 阶复矩阵 A, 都有

$$\log|\lambda(A)| \prec \log\sigma(A). \tag{8.7.8}$$

证明 注意到

$$\begin{pmatrix} (AA^*)^{1/2} & A \\ A^* & (A^*A)^{1/2} \end{pmatrix} \geqslant 0,$$

推论证毕.

注 由 $\log x \prec_w \log y$ 可推出 $x \prec_w y$, 见 Bhatia(1997, p.42.) 因此, 本节得到的关于 Log- 弱受控的不等式在弱受控下也成立.

§8.8 随机矩阵

一个以随机变量为元素的矩阵称为随机矩阵. 如果它只有一行或一列, 则称为随机向量. 本节讨论与随机矩阵有关的一些受控关系.

在涉及随机性的自然科学及社会科学领域中, 一类重要的随机矩阵来自于随机向量的协方差阵和相关阵的估计, 假设 X 为 $p \times 1$ 的随机向量, 均值 $\mu = EX$. X 的方差 —— 协方差阵 (通常简称为协方差阵) 和相关阵分别定义为

$$\operatorname{cov}(X) = E(X-\mu)(X-\mu)' \triangleq \Sigma = (\sigma_{ij})_{p\times p},$$

$$\Gamma = (r_{ij}), \qquad \text{其中 } r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

当我们有了样本 $X_1, \cdots, X_n$ 之后, 就可以对 Σ 和 Γ 作出如下估计

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=j}^{n}(X_i - \overline{X})(X_i - \overline{X})' = (\hat{\sigma}_{ij}),$$

$$\hat{\Gamma} = (\hat{r}_{ij}),$$

其中

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}.$$

容易看到, $\hat{\Sigma}$ 和 $\hat{\Gamma}$ 都是实对称矩阵, 在应用上, 最重要的情形是, $X_1, \cdots, X_n$ 为来自 p 维正态总体 $N(\mu, \Sigma)$ 的简单随机样本, 此时, $\overline{X}$, $\hat{\Sigma}$ 和 $\hat{\Gamma}$ 分别为 μ, Σ 和 Γ 的极大似然估计, 且当 $n > p$ 时 $\hat{\Sigma}$ 和 $\hat{\Gamma}$ 为正定阵 (以概率为 1). 于是随机阵 $\hat{\Sigma}$ 和 $\hat{\Gamma}$ 为正定的实对称阵. 在多元统计分析中我们对这样一类随机阵的特征值往往特别感兴趣.

本节的讨论略一般一些, 我们将研究正定的 Hermite 随机阵的特征值. 先证明几个引理.

引理 8.8.1(Jensen 不等式) 设 $X_1, \cdots, X_n$ 为随机变量, 均值 $EX_i = \mu_i$ $(i = 1, \cdots, n)$. 则对一切连续凸函数 ϕ, 有

$$\phi(\mu_1, \cdots, \mu_n) \leqslant E\phi(X_1, \cdots, X_n). \tag{8.8.1}$$

证明 因 ϕ 为连续凸函数, 故对任意固定的 $X_0 \in R^n$ 和一切 $X \in R^n$, 有

$$\phi(X) \geqslant \phi(X_0) + c'(X - X_0), \tag{8.8.2}$$

其中 c 为不依赖于 X 的 $n \times 1$ 常数向量 (对 $n = 1$ 的情况, 证明见 Rao(1973), 对一般的 n 可类似地证明). 现取 $X_0 = \mu = (\mu_1, \cdots, \mu_n)'$. 并对 (8.8.2) 两边求均值, 注意到 $E(X - X_0) = E(X - \mu) = 0$. 便得 (8.8.1). 证毕.

引理 8.8.2 设 A 为 R^m 的凸子集, X 为取值在 A 上的随机向量, 且 $\mu = EX$ 存在, 又 $f_1, \cdots, f_n$ 为定义在 A 上的连续函数, 满足

(1) $f_1(\mu) \geqslant \cdots \geqslant f_n(\mu)$;

(2) 对 $k = 1, \cdots, n, \sum\limits_{i=1}^{k} f_i(X)$ 为 A 上的凸函数, 则

$$(f_1(\mu), \cdots, f_n(\mu))' \prec_w (Ef_1(X), \cdots, Ef_n(X))',$$

这里假定所涉及的均值都存在.

证明 首先注意到, 由 A 的凸性, 可推出 $\mu \in A$. 因为 $\sum\limits_{i=1}^{k} f_i(X)$ 为连续凸函数, 应用 Jensen 不等式, 得

$$\sum_{i=1}^{k} Ef_i(X) = E\sum_{i=1}^{k} f_i(X) \geqslant \sum_{i=1}^{k} f_i(\mu),$$

再结合假设条件 (1), 引理得证.

引理 8.8.3 设 $A \subset R^m$, $f_1, \cdots, f_n$ 定义

$$\sum_{i=1}^{k} f_i(x) = \sup_{u \in S_k} g(u, x), \qquad k = 1, \cdots, n, \quad x \in A,$$

其中 $S_1, \cdots, S_n$ 为 n 个集合, 使得 $f_1(x) \geqslant \cdots \geqslant f_n(x)$ 对一切 $x \in A$ 成立. 若对每个固定的 $u \in \bigcup\limits_{k=1}^{n} S_k$, $g(u, x)$ 为 x 的连续凸函数, 则对一切取值在 A 上的随机向量 X, $\mu = EX \in A$, 有

$$(f_1(\mu), \cdots, f_n(\mu))' \prec_w (Ef_1(X), \cdots, Ef_n(X))'.$$

证明 因为对每个固定的 $u \in \bigcup\limits_{k=1}^{n} S_k$, 函数

$$\sum_{i=1}^{k} f_i(x) = \sup_{u \in S_k} g(u, x)$$

是 x 的连续凸函数, 且 $f_1(\mu) \geqslant \cdots \geqslant f_n(\mu)$, 利用上一个引理, 立得所需的结论. 证毕.

对于一个 $n \times n$ 的方阵 A, 假定它的特征值都是实数, 那么和以前一样, 我们总把它的特征值作递减排列, 即 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$. 现在我们来研究 Hermite 随机阵的特征向量的一些受控关系.

定理 8.8.1 设 X 为 $n \times n$ Hermite 随机阵, 则

$$(\lambda_1(EX), \cdots, \lambda_n(EX))' \prec (E\lambda_1(X), \cdots, E\lambda_n(X))'. \tag{8.8.3}$$

证明 记 $S_k = \{U\text{: } U \text{ 为 } k \times n \text{ 复阵}, UU^* = I_k\}$. 由定理 6.6.1, 得

$$\sum_{i=1}^{k} \lambda_i(EX) = \max_{U \in S_k} \mathrm{tr} U(EX)U^*, \quad k = 1, \cdots, n. \tag{8.8.4}$$

视 $\lambda_i(EX)$ 为引理 8.8.3 中的 f_i, 容易看出, 它们满足引理 8.8.3 的条件, 这就证明了弱受控关系. 再结合

$$\sum_{i=1}^{n} E\lambda_i(X) = \mathrm{tr} E(X) = \sum_{i=1}^{n} \lambda_i(EX),$$

便证明了 (8.8.3). 证毕.

推论 8.8.1 设 X 为 $n \times n$ Hermite 随机阵, EX 存在, 则

$$E\lambda_1(X) \geqslant \lambda_1(EX),$$
$$E\lambda_n(X) \leqslant \lambda_n(EX).$$

当 X 为 Hermite 阵时, $EX = (\mu_{ij})$ 也是 Hermite 阵, 应用定理 8.3.1, 得

推论 8.8.2 设 X 为 Hermite 随机阵, 记 $EX = (\mu_{ij})$, 则

$$(\mu_{11}, \cdots, \mu_{nn})' \prec (\lambda_1(EX), \cdots, \lambda_n(EX))'.$$

定理 8.8.2 设 X 为 $m \times m$ 正定 Hermite 随机阵, $\Theta = EX$, A 为 $n \times m$ 矩阵. 则

$$(\lambda_1(A\Theta^{-1}A^*), \cdots, \lambda_n(A\Theta^{-1}A^*))'$$
$$\prec_w (E\lambda_1(AX^{-1}A^*), \cdots, E\lambda_n(AX^{-1}A^*))'$$

证明 设 S_k 定义同定理 8.8.1. 类似于定理 8.8.1 的证明, 我们有

$$\sum_{i=1}^{k} \lambda_i(A\Theta^{-1}A^*) = \max_{U \in S_k} \mathrm{tr} UA\Theta^{-1}A^*U^*, \qquad k = 1, \cdots, n,$$

对每个固定的 A 和 U, 上式等号右边为 Θ 的凸函数 (见 Marshall-Olkin(1979)), 由引理 8.8.3, 定理得证.

假设 X_1 和 X_2 为两个同阶实对称正定的随机阵, $EX_i = \Theta_i$, 在多元统计分析中, 我们需要研究方程 $|X_1 - \lambda X_2| = 0$ 的根 (见王学仁, 王松桂 (1990)), 也就是随机阵 $X_1X_2^{-1}$ 的特征值, 下面的定理建立了有关这样随机阵的特征值的受控关系.

定理 8.8.3 设 X_1 和 X_2 为两个独立的 $n\times n$ 正定 Hermite 随机阵, $EX_i = \Theta_i$, $i = 1, 2$. 则

$$
\begin{aligned}
&(\lambda_2(\Theta_2\Theta_1^{-1}), \cdots, \lambda_n(\Theta_2\Theta_1^{-1}))' \\
&\prec_w (E\lambda_1(X_2X_1)^{-1}, \cdots, E\lambda_n(X_2X_1^{-1}))'.
\end{aligned}
$$

证明 应用 X_1 和 X_2 的独立性, 两次应用定理 8.8.2, 得

$$
\begin{aligned}
&(\lambda_1(\Theta_2\Theta_1^{-1}), \cdots, \lambda_n(\Theta_2\Theta_1^{-1}))' \\
=&(\lambda_1(\Theta_2^{1/2}\Theta_1^{-1}\Theta_2^{1/2}), \cdots, \lambda_n(\Theta_2^{1/2}\Theta_1^{-1}\Theta_2^{1/2}))' \\
\prec_w&(E_1\lambda_1(\Theta_2^{1/2}X_1^{-1}\Theta_2^{1/2}), \cdots, E_1\lambda_n(\Theta_2^{1/2}X_1^{-1}\Theta_2^{1/2}))' \\
=&(E_1\lambda_1(X_1^{-1/2}\Theta_2X_1^{-1/2}), \cdots, E_1\lambda_n(X_1^{-1/2}\Theta_2X_1^{-1/2}))' \\
\prec_w&(E\lambda_1(X_1^{-1/2}X_2X_1^{-1/2}), \cdots, E\lambda_n(X_1^{-1/2}X_2X_1^{-1/2}))' \\
=&(E\lambda_1(X_2X_1^{-1}), \cdots, E\lambda_n(X_2X_1^{-1}))'.
\end{aligned}
$$

其中 E_1 是对 X_1 求均值.

§8.9 复合矩阵

在本章的几个定理的证明中, 我们用到了复合矩阵的概念, 鉴于这个概念用处较少, 且局限于这一章, 所以为了读者阅读方便, 我们把复合阵的一些基本结果放在本节内.

定义 8.9.1 设 A 为 $m\times n$ 复矩阵, $1 \leqslant k \leqslant \min(m, n)$. A 的第 k 次复合是一个 $\begin{pmatrix} m \\ k \end{pmatrix} \times \begin{pmatrix} n \\ k \end{pmatrix}$ 的矩阵, 它的元素为

$$
\det A\begin{pmatrix} i_1\cdots i_k \\ j_1\cdots j_k \end{pmatrix},
$$

这些元素依 $i_1, \cdots, i_k, j_1, \cdots, j_k$ 按字典序排列. A 的第 k 次复合阵记为 $A^{(k)}$.

根据字典序排列法则, 在 $A^{(k)}$ 中两个元素

$$a = \det A\begin{pmatrix} i_1\cdots i_k \\ j_1\cdots j_k \end{pmatrix}, \quad b = \det A\begin{pmatrix} l_1\cdots l_k \\ t_1\cdots t_k \end{pmatrix}.$$

若 $(i_1,\cdots,i_k,j_1,\cdots,j_h)$ 与 $(l_1,\cdots,l_h,t_1,\cdots,t_k)$ 的第一个不相等的对应分量, 前者小, 则将 a 排在 b 的前面, 否则, 将 b 排在 a 的前面.

根据定义容易验证, $A^{(k)}$ 具有下列性质:

(1) $A^{(1)} = A$, $A^{(n)} = \det A$;

(2) $(A^{(k)})^* = (A^*)^{(k)}$, $(A^{(k)})' = (A')^{(k)}$;

(3) 若 A 是 Hermite 阵, 则 $A^{(k)}$ 也是 Hermite 阵;

(4) 若 A 是上 (下) 三角阵, 则 $A^{(k)}$ 也是上 (下) 三角阵.

定理 8.9.1　设 A 和 B 分别为 $p\times m$ 和 $m\times n$ 复矩阵, $1\leqslant k\leqslant \min(p,m,n)$. 则 $(AB)^{(k)} = A^{(k)}B^{(k)}$.

证明见 Muir(1906).

推论 8.9.1　设 A 为可逆阵, 则 $(A^{-1})^{(k)} = (A^{(k)})^{-1}$.

证明　对 $AA^{-1} = I$ 应用定理, 得到 $A^{(k)}(A^{-1})^{(k)} = I^{(k)} = I$, 即 $A^{(k)}$ 的逆为 $(A^{-1})^{(k)}$.

这个事实表明, 求逆运算和求复合阵的运算可以互换次序.

推论 8.9.2　设 U 为 $n\times n$ 酉阵. 则 $U^{(k)}$ 仍为酉阵.

证明　$(U^{(k)})^* = (U^*)^{(k)} = (U^{-1})^{(k)} = (U^{(k)})^{-1}$.

定理 8.9.2　设 A 为 $n\times n$ 复方阵, 特征值为 $\lambda_1,\cdots,\lambda_n$. 则 $A^{(k)}$ 的特征值为

$$\prod_{j=1}^{k}\lambda_{i_j}, \quad 其中,\ 1\leqslant i_1<\cdots<i_k\leqslant n.$$

证明　由定理 1.5.7, 存在酉阵 U 和下三角阵 L, 使得 $A = UL\times L^*$, L 的对角元为 A 的特征值. 根据前面叙述的 $A^{(k)}$ 的性质, 有

$$A^{(k)} = (ULU^*)^{(k)} = U^{(k)}L^{(k)}(U^*)^{(k)} = U^{(k)}L^{(k)}(U^{(k)})^*.$$

注意到 $L^{(k)}$ 仍为下三角阵, 它的对角元为 $\prod_{j=1}^{k}\lambda_{i_j}$ 其中 $1\leqslant i_1<\cdots<i_k\leqslant n$, $U^{(k)}$ 为酉阵, 所以 $A^{(k)}$ 的特征值为 $L^{(k)}$ 的对角元. 定理证毕.

推论 8.9.3　若 A 为半正定 (正定)Hermite 阵, 则 $A^{(k)}$ 也是半正定 (正定)Hermite 阵.

证明 若 A 是 $n\times n$ 半正定 (正定)Hermite 阵, $\lambda_1\geqslant\cdots\geqslant\lambda_n\geqslant 0$ 为其特征值, 则 $A^{(k)}$ 的最小特征值为

$$\prod_{1=i}^{k}\lambda_{n-i+1}(A)\geqslant 0.$$

因 $A^{(k)}$ 也是 Hermite 阵, 所以 $A^{(k)}$ 是半正定 (正定)Hermite 阵.

定理 8.9.3 设 A 为 $n\times n$ 复方阵, $\lambda_1,\cdots,\lambda_n$ 为其特征值, 则

$$\mathrm{tr}A^{(k)}=S_k(\lambda_1,\cdots,\lambda_n),$$

其中 $S_k(\lambda_1,\cdots,\lambda_n)$ 为 $\lambda_1,\cdots,\lambda_n$ 的初等对称函数 (其定义见 §8.2).

结论由 S_k 的定义及定理 8.9.2 立即推得.

第 9 章　在线性统计中的若干应用举例

我们知道, 数学具有两大特点, 即结论的抽象性和应用的广泛性, 前者是后者的基础, 而后者是前者的必然结果. 本书所讨论的有关矩阵的不等式也不例外, 关于它们的抽象性读者从前面的讨论已经清楚地看到了. 本章的目的, 仅仅就其应用做一些示例性的解释. 因为矩阵不等式已广泛地应用于自然科学和工程技术的很多领域, 限于篇幅与作者的水平, 我们的讨论仅局限于在数理统计中的若干应用.

本章前两节将介绍矩阵偏序和受控在线性统计参数估计中的应用, 在第三节, 我们应用一个统计定理把 Kantorovich 不等式推广到带约束的情形; 最后一节介绍受控在统计检验方面的应用.

§9.1　估计与模型的比较

我们先引进一般线性统计模型.

假设因变量 Y 和自变量 $X_1,\cdots,X_{p-1}$ 之间有线性关系

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+\varepsilon,$$

其中 $\beta_0,\beta_1,\cdots,\beta_{p-1}$ 为未知参数, ε 为随机误差. 为了建立这个模型, 我们首先需要基于观测数据对未知参数 $\beta_0,\beta_1,\cdots,\beta_{p-1}$ 作出估计. 现在假定对 p 个变量 Y, $X_1,\cdots,X_{p-1}$ 作了 n 次观测: $y_i,x_{i1},\cdots,x_{ip-1}$, $i=1,\cdots,n$, 它们满足

$$y_i=\beta_0+x_{i1}\beta_1+\cdots+x_{ip-1}\beta_{p-1}+\varepsilon_i,\qquad i=1,\cdots,n,\tag{9.1.1}$$

ε_i 为第 i 次试验的随机误差, 其均值 $E(\varepsilon_i)=0$, 协方差 $\mathrm{cov}(\varepsilon_i,\varepsilon_j)=\sigma^2v_{ij}$, $i,j=1,\cdots,n$. 若引进矩阵记号

$$y=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix},\quad X=\begin{pmatrix}x_{10}&x_{11}&\cdots&x_{1,p-1}\\ x_{20}&x_{21}&\cdots&x_{2,p-1}\\ \vdots&\vdots&&\vdots\\ x_{n0}&x_{n1}&\cdots&x_{n,p-1}\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{p-1}\end{pmatrix},\quad \varepsilon=\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{pmatrix},$$

其中 $x_{i0}=1$, $i=1,\cdots,n$. 则 (9.1.1) 变为

$$x=X\beta+\varepsilon,\quad E(\varepsilon)=0,\quad \mathrm{cov}(\varepsilon)=\sigma^2V,\tag{9.1.2}$$

这里 $V=(v_{ij})$. X 被称为设计阵, 其秩 $r(X)=t\leqslant p$. 关于协方差阵, 一般性的假设是 $V\geqslant 0$, 即 V 为实对称半正定阵, 但最常见的假设是 $V>0$, 且已知. 如果我们有充分理由认为在不同次试验中, 随机误差 ε_i 和 ε_j 是等方差且互不相关, 也就是 $v_{ij}=0$, 当 $i\neq j$ 时; 而当 $i=j$ 时, $v_{11}=\cdots=v_{nn}$, $V=I$. 在统计文献中, 这称为 Gauss-Markov 假设.

在统计学中, 我们最重要的兴趣是针对设计阵 X 和协方差阵 V 的不同情况, 对 β 或它的线性组合 $c'\beta$(这里 c 为 $p\times 1$ 常数向量) 作出具有良好性质的估计, 所采用的方法主要是最小二乘原理. 为了下面讨论方便, 我们把这方面的重要结果概括如下 (详细讨论见王松桂 (1987) 或 Wang Songgui 等 (1993)).

(1) $V=I$

若 $r(X)=t<p$, 则对任意 $c\in\mathscr{M}(X')$, $c'\beta$ 称为可估函数. 记

$$\hat{\beta}=(X'X)-X'y, \tag{9.1.3}$$

则对一切可作函数 $c'\beta$, $c'\hat{\beta}$ 与广义逆 $(X'X)^-$ 的选择无关, 且 $\hat{c}'\beta$ 为 $c'\beta$ 的唯一最佳线性无偏估计 (Best Linear Unbiased Estimator, 简记为 BLU 估计). 当 $r(X)=p$ 时, $X'X$ 为可逆阵, 此时 β 本身是可估的, 我们称 $\hat{\beta}=(X'X)^{-1}X'y$ 为 β 的最小二乘估计 (Least Square estimator, 简记为 LS 估计), 它也是 β 的 BLU 估计.

(2) $V>0(V\neq I)$ 且已知

当 $r(X)=t<p$ 时, 对一切可估函数 $c'\beta$, $c'\beta^*$ 为其唯一 BLU 估计, 这里

$$\beta^*=(X'V^{-1}X)^-X'V^{-1}y. \tag{9.1.4}$$

当 $r(X)=p$ 时, β 是可估的, $\beta^*=(X'V^{-1}X)^{-1}X'V^{-1}y$ 称为 β 的广义最小二乘估计 (Generalized LS 估计, 简记为 GLS 估计).

(3) $V\geqslant 0$ 且已知

记 $T=V+XUX'$, 这里 U 为实对称的半正定阵, 使得 $r(T)=r(V\,\vdots\,X)$. 对一切可估函数 $c'\beta$, $c'\beta^*$ 为其唯一的 BLU 估计, 这里

$$\beta^*=(X'T^-X)^-X'T^-y. \tag{9.1.5}$$

有了这些准备之后, 我们来讨论矩阵不等式在估计比较中的应用.

在统计学中, 对不同估计进行比较的准则很多, 如果被比较的估计都是无偏估计, 那么协方差阵是最常用的标准. 现在我们暂时离开线性统计模型 (9.1.2), 考虑一般的参数估计问题. 假设 $\tilde{\theta}_1$ 和 $\tilde{\theta}_2$ 为未知参数向量 θ 的两个无偏估计, 当它们的协方差阵满足

$$\operatorname{cov}(\tilde{\theta}_1)\geqslant\operatorname{cov}(\tilde{\theta}_2) \tag{9.1.6}$$

时, 则称 $\tilde{\theta}_2$ 至少与 $\tilde{\theta}_1$ 一样好. 假若 $\tilde{\theta}_1$ 和 $\tilde{\theta}_2$ 中至少有一个是无偏估计, 则采用均方误差矩阵 (Mean Square Error Matrix, 简记为 MSEM) 比较合理, 它定义为

$$\mathrm{MSEM}(\tilde{\theta}_1)=E(\tilde{\theta}_1-\theta)(\tilde{\theta}_1-\theta)'.$$

如果

$$\mathrm{MSEM}(\tilde{\theta}_1)\geqslant \mathrm{MSEM}(\tilde{\theta}_2), \tag{9.1.7}$$

则称在均方误差阵意义下 $\tilde{\theta}_2$ 至少与 $\tilde{\theta}_1$ 一样好.

注意到 (9.1.6) 和 (9.1.7) 都是矩阵的偏序关系, 这表明, 协方差阵和均方误差阵的偏序刻画了估计的优劣.

现在回到线性统计模型 (9.1.2). 考虑 $V>0$, $r(X)=p$ 的情形, 根据前面叙述的结果, 我们有

$$\mathrm{cov}(\hat{\beta})\geqslant \mathrm{cov}(\beta^*),$$

即

$$(X'X)^{-1}X'VX(X'X)^{-1}\geqslant (X'V^{-1}X)^{-1},$$

这个偏序关系就是 β^* 优于 $\hat{\beta}$ 的一个刻画.

对于 MSEM 准则, 为简单计, 设 $V=I$, 也就是 Gauss-Markov 假设成立. 设 $\tilde{\beta}_i=A_iy$, $i=1,2$ 为 β 的两个有偏估计, 则它们的均方误差阵为

$$\mathrm{MSEM}(\tilde{\beta}_i)=\sigma^2A_iA_i'+b_ib_i', \qquad i=1,2,$$

其中 $b_i=(A_iX-I)\beta$ 称为估计 $\tilde{\beta}_i$ 的偏差向量. 记

$$D=\sigma^2[A_1A_1'-A_2A_2'], \tag{9.1.8}$$

则它们的均方误差矩阵之差可表为

$$\mathrm{MSEM}(\tilde{\beta}_1)-\mathrm{MSEM}(\tilde{\beta}_2)=D+b_1b_1'-b_2b_2'.$$

为了保证 $\tilde{\beta}_2$ 至少与 $\tilde{\beta}_1$ 一样好, 我们就要求上式右边为一个半正定阵, 应用推论 7.2.4, 我们立即得到

定理 9.1.1 在线性统计模型 (9.1.2) 中, 设 $V=I$, $r(X)=p$, 则 β 的线性估计 $\tilde{\beta}_2$ 至少与 $\tilde{\beta}_1$ 一样好, 当且仅当下列三个条件成立.

(1) $D+b_1b_1'\geqslant 0$;

(2) $b_2\in\mathscr{M}(D+b_1b_1')$;

(3) $b_2'(D+b_1b_1')^-b_2\leqslant 1$.

条件 (1) 的一个充分条件是 $D \geqslant 0$. 利用定理 7.2.4, 我们得到 $D \geqslant 0$ 的一个等价条件为 $\mathscr{M}(A_2) \subset \mathscr{M}(A_1)$, $\lambda_1(A_2'(A_1A_1') + A_2) \leqslant 1$, 这里 $\lambda_1(A)$ 表示 A 的最大特征值. 据此我们可以得到如下推论.

推论 9.1.1 在定理 9.1.1 假设下, β 的线性估计 Ay 满足 $\text{cov}(Ay) \leqslant \text{cov}(\hat{\beta})$ 当且仅当 $\lambda_1(A'X'XA) \leqslant 1$.

关于这方面的详细讨论请参阅 Baksalary 等 (1989).

下面两个例子是模型 (9.1.2) 的特殊情形, 它们显示了受控理论在估计比较中的应用.

例 9.1.1 考虑最简单的模型

$$y_i = \mu + \varepsilon_i, \qquad i = 1, \cdots, n,$$

其中 $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$, $\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2\rho$, $i \neq j$.

如果 $w_i \geqslant 0$, $i = 1, \cdots, n$ 满足 $\sum_{i=1}^{n} w_i = 1$, 那么

$$\tilde{\mu} = \sum_{i=1}^{n} w_i y_i$$

为 μ 的线性无偏估计. 它的方差

$$\begin{aligned}
\text{var}(\tilde{\mu}) &= \sigma^2 \sum_{i=1}^{n} w_i^2 + \sigma^2\rho \sum_{i \neq j} w_i w_j \\
&= \sigma^2 \sum_{i=1}^{n} w_i^2 + \sigma^2\rho \left[\left(\sum_{i=1}^{n} w_i \right)^2 - \sum_{i=1}^{n} w_i^2 \right] \\
&= \sigma^2(1-\rho) \sum_{i=1}^{n} w_i^2 + \sigma^2\rho.
\end{aligned}$$

由例 8.1.2 知, 对一切满足 $w_i \geqslant 0$, $\Sigma w_i = 1$ 的 $w_1, \cdots, w_n$, 有

$$\left(\frac{1}{n}, \cdots, \frac{1}{n} \right)' < (w_1, \cdots, w_n)'.$$

再利用推论 8.2.3, 可推出 $\sum_{i=1}^{n} w_i^2$ 为 Schur 凸函数. 因此

$$\mu^* = \frac{1}{n} \sum_{i=1}^{n} y_i$$

在无偏估计类 $\tilde{\mu}=\sum_{i=1}^{n}w_iy_i$, $w_i\geqslant 0$, $\sum_{i=1}^{n}w_i=1$ 中具有最小方差. 容易看到, 去掉条件 $w_i\geqslant 0$, 这个结论仍然成立, 于是 μ^* 在一切线性无偏估计中具有最小方差, 故它是 BLU 估计. 请注意, μ^* 并不依赖于 ρ, 因此 μ^* 的最优性并不能从 Gauss-Markov 定理直接推出. 证明 μ^* 的这种最优性的另一种方法是 LS 估计的稳健性 (详见王松桂 (1987, p. 184)).

例 9.1.2　单向分类随机模型

$$y_{ij}=\mu+\alpha_i+\varepsilon_{ij},\qquad i=1,\cdots,k;j=1,\cdots,n_i,$$

其中 α_i 为随机效应, ε_{ij} 为随机误差, 假定它们都相互独立, 均值为零, $\mathrm{var}(\varepsilon_{ij})=\sigma_\varepsilon^2$, $\mathrm{var}(\alpha_i)=\sigma_\alpha^2$. 假设 $n_1\geqslant\cdots\geqslant n_k$, $\sum_{i=1}^{k}n_i=n$.

我们考虑总平均 μ 的如下估计

$$\tilde{\mu}=\sum_{i=1}^{k}w_i\bar{y}_{i.},$$

其中 $\sum_{i=1}^{n}w_i=1$, $\bar{y}_{i.}=\frac{1}{n}\sum_{j=1}^{n_i}y_{ij}$.

将 $\tilde{\mu}$ 改写为

$$\tilde{\mu}=\mu+\sum_{i=1}^{k}w_i\alpha_i+\sum_{i=1}^{k}w_i\bar{\varepsilon}_i,$$

这里

$$\bar{\varepsilon}_{i.}=\frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij},\qquad i=1,\cdots,k.$$

$\tilde{\mu}$ 是 μ 的无偏估计, 它的方差为

$$\begin{aligned}\mathrm{var}(\tilde{\mu})=&\sigma_\alpha^2\sum_{i=1}^{k}w_i^2+\sigma_e^2\sum_{i=1}^{k}w_i^2/n_i\\ \triangleq&\sigma_\alpha^2V_\alpha(\tilde{\mu})+\sigma_e^2V_e(\tilde{\mu}),\end{aligned}$$

这里

$$V_\alpha(\tilde{\mu})=\sum_{i=1}^{k}w_i^2,\quad V_e(\tilde{\mu})=\sum_{i=1}^{k}w_i^2/n_i.$$

对于 w_i 的如下三种选择:

(a) $w_i = \dfrac{1}{k}$;

(b) $w_i = \dfrac{n_i}{n}$;

(c) $w_i = \dfrac{n_i(n-n_i)}{n^2 - \Sigma n_j^2}$.

我们把对应的三种估计分别记为 $\tilde{\mu}_1$, $\tilde{\mu}_2$ 和 $\tilde{\mu}_3$. 则有

$$V_\alpha(\tilde{\mu_1}) \leqslant V_\alpha(\tilde{\mu}_3) \leqslant V_\alpha(\tilde{\mu}_2); \tag{9.1.9}$$

$$V_e(\tilde{\mu}_1) \leqslant V_e(\tilde{\mu}_3) \leqslant V_e(\tilde{\mu}_2). \tag{9.1.10}$$

这些不等式是下列两条事实的直接推论.

(a) $\sum\limits_{i=1}^{k} w_i^2$ 和 $\sum\limits_{i=1}^{k} w_i^2/n_i$ 是 Schur 凸函数;

(b) $$\left(\frac{1}{k},\cdots,\frac{1}{k}\right)' \prec \left(\frac{n_1(n-n_1)}{n_2-\Sigma n_j^2},\cdots,\frac{n_k(n-n_k)}{n^2-\Sigma n_j^2}\right)'$$

$$\prec \left(\frac{n_1}{n},\cdots,\frac{n_k}{n}\right)'.$$

(a) 可以直接从推论 8.2.3 导出. 下面证明 (b). 在 (b) 中, 左边受控关系可由例 8.1.2 得到. 为证右边的受控关系, 不妨假设 $n_1 \geqslant \cdots \geqslant n_k$. 因为由 $x_1 \geqslant x_2$ 和 $x_1 + x_2 \leqslant 1$ 可以证明 $x_1(1-x_1) \leqslant x_2(1-x_2)$. 故有

$$n_1(n-n_1) \geqslant \cdots \geqslant n_k(n-n_k).$$

若记

$$x_i = \frac{n_1(n-n_i)}{n^2-\Sigma n_j^2}, \qquad y_i = \frac{n_i}{n},$$

则

$$\frac{y_i}{x_i} = \frac{n^2-\Sigma n_j^2}{(n-n_i)n}, \qquad i = 1,\cdots,k.$$

它关于 $i = 1,\cdots,k$ 是单调减的. 应用定理 8.1.10 立得 $(x_1,\cdots,x_k)' \prec (y_i,\cdots,y_k)'$. 这就证明了 (b).

综合 (9.1.9) 和 (9.1.10), 我们得到如下结论

$$\operatorname{var}(\tilde{\mu}_1) \leqslant \operatorname{var}(\tilde{\mu}_3) \leqslant \operatorname{var}(\tilde{\mu}_2).$$

现在我们转入本节第二个问题 —— 模型比较的讨论.

为了符号简单计, 我们把线性统计模型 (9.1.2) 记为 $d = L(y, X\beta, \sigma^2 V)$. 一个模型就代表了一组试验. 当 $r(X) = p$, $r(V) = n$ 时, 从这组试验我们可以得到 β 的

BLU 估计 $\beta^* = (X'V^{-1}X)^{-1}\ X'V^{-1}y$, 它的协方差阵 $\mathrm{cov}(\beta^*) = \sigma^2(X'V^{-1}X)^{-1}$. 如果我们有两组试验 $d_i = L(y_i, X_i\beta, \sigma^2 V_i)$, $i = 1, 2$. 那么从这两组数据就可获得 β 的两个 BLU 估计

$$\beta_i^* = (X_i'V_i^{-1}X_i)^{-1}X_i'V_i^{-1}y_i, \qquad i = 1, 2.$$

如果

$$\mathrm{cov}(\beta_1^*) \leqslant \mathrm{cov}(\beta_2^*), \tag{9.1.11}$$

则从试验 $d_1 = L(y_1, X_1\beta, \sigma^2 V_1)$ 得到的 β 的 BLU 估计至少与从试验 $d_2 = L(y_2, X_2\beta, \sigma^2 V_2)$ 得到的 β 的 BLU 估计一样好, 此时称 d_1 至少与 d_2 一样好, 记为 $d_1 \geqslant d_2$. 据此, 我们就可以对两组试验进行比较.

根据上面的定义, 知

$$\begin{aligned} d_1 \geqslant d_2 &\Longleftrightarrow (X_1'V_1^{-1}X_1)^{-1} \leqslant (X_2'V_2^{-1}X_2)^{-1} \\ &\Longleftrightarrow X_1'V_1^{-1}X_1 \geqslant X_2'V_2^{-1}X_2, \end{aligned} \tag{9.1.12}$$

这里应用了推论 7.2.2. 从 (9.1.12) 看出, 当我们比较两个试验时, 不需要考虑观测向量 y_i $(i = 1, 2)$, 只需比较两组试验条件, 它们集中反映在 X_i 和 V_i 上.

上面考虑的是一个简单情况, 即 $r(X_i) = p$. 但往往设计阵是列降秩的, 即 $r(X_i) < p$. 此时为了比较两组试验, 我们需要比较所有可估函数的 BLU 估计的方差. 记 $\beta_i^* = (X_i'V_i^{-1}X_i)^- X_i'V_i^{-1}y_i$, $i = 1, 2$. 如果在 d_2 中任一可估函数 $c'\beta$, 在 d_1 也是可估计的, 并且

$$\mathrm{var}(c'\beta_1^*) \leqslant \mathrm{var}(c'\beta_2^*),$$

则称 d_1 至少与 d_2 一样好, 记为 $d_1 \geqslant d_2$.

前面我们对 $r(x_i) = p$ 的情形, 证明了 $d_1 \geqslant d_2$ 当且仅当 (9.1.12) 成立, 下面的定理表明, 同样的结论对 $r(X_i) < p$ 的情形也成立.

定理 9.1.2　对于两个线性统计模型 $d_i = L(y_i, X_i\beta, \sigma^2 V_i)$, $i = 1, 2$. 其中 $V_i > 0$, $i = 1, 2$, 则 $d_1 \geqslant d_2$ 当且仅当

$$X_1'V_1^{-1}X_1 \geqslant X_2'V_2^{-1}X_2. \tag{9.1.13}$$

证明　依定义, 得

$$d_1 \geqslant d_2 \Longleftrightarrow \begin{cases} \mathscr{M}(X_2') \subset \mathscr{M}(X_1'), \\ \mathrm{var}(c'\beta_1^*) \leqslant \mathrm{var}(c'\beta_2^*), \quad \text{对任意 } c \in \mathscr{M}(X_2'). \end{cases}$$

$$\Longleftrightarrow \begin{cases} \mathscr{M}(X_2') \subset \mathscr{M}(X_1'), \\ \text{对一切 } c \in \mathscr{M}(X_2'),\ c'(X_1'V_1^{-1}X_1)^- c \leqslant c'(X_2'V_2^{-1}X_2)^- c. \end{cases}$$

记 $Z_i = X_i'V^{-1/2}$, $A_i = Z_i'Z_i$, 并利用 $\mathscr{M}(X_2') = \mathscr{M}(Z_2'Z_2) = \mathscr{M}(A_2)$, 知

$$\text{上式} \Longleftrightarrow \begin{cases} \mathscr{M}(A_2) \subset \mathscr{M}(A_1) \\ A_2A_1^-A_2 \leqslant A_2 \end{cases} \Longleftrightarrow A_1 \geqslant A_2, \text{此即 (9.1.13). 证毕.}$$

上面的结果还可以进一步推广到 $V_i \geqslant 0$ 的情况 (详见 Stepniak Wang 和 Wu(1984)).

§9.2 相对效率

我们沿用上节的记号, 对线性统计模型 (9.1.2), 当 $V > 0$, $r(X) = p$ 时, β 的 GLS 估计 β^* 也是 BLU 估计. 但在应用上, V 往往是未知的, 于是人们经常采用 LS 估计 $\hat{\beta} = (X'X)^{-1}X'y$ 来代替 β^*. 这种替代所产生的信息损失可以用这两种估计的广义方差比

$$\frac{\det(\text{cov}(\beta^*))}{\det(\text{cov}(\hat{\beta}))} \triangleq \text{RE}(\hat{\beta})$$

来度量, 称其为 $\hat{\beta}$ 关于 β^* 的相对效率, 显然, $0 \leqslant \text{RE}(\hat{\beta}) \leqslant 1$. 问题是求 $\text{Re}(\hat{\beta})$ 的下界, 它刻画了用 $\hat{\beta}$ 代替 β^* 所产生的最大信息损失.

本节的目的是介绍矩阵不等式在求 $\text{RE}(\hat{\beta})$ 下界方面的应用.

定理 9.2.1 在线性统计模型 (9.1.2) 中, 设 $V > 0, \lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 V 的特征值, $r(X) = p$, $n \geqslant 2p$. 则

$$\text{RE}(\hat{\beta}) \geqslant \prod_{i=1}^{p} \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}. \tag{9.2.1}$$

证明 因为

$$\text{RE}(\hat{\beta}) = [\det(X'X)^{-1}X'VX(X'X)^{-1}\det(X'V^{-1}X)]^{-1} \tag{9.2.2}$$

在 X 右乘可逆阵之后保持不变, 故应用 Schmidt 正交化程序, 我们可以假设在 (9.2.2) 中, $X'X = I$. 于是

$$\text{RE}(\hat{\beta}) = (\det(X'VX)\det(X'V^{-1}X))^{-1},$$

再应用定理 4.10.3, 便得到 (9.2.1). 证毕

当 $r(X) \leqslant p$ 时, β 可能是不可估的, 此时我们需考虑可估函数 $c'\beta$ 的估计. 与前面的讨论相类似, 当 V 未知时, 我们需要用 $c'\beta$ 的 LS 估计 $c'\hat{\beta}$ 代替它的 BLU 估计 $c'\beta^*$. 这种代替所产生的信息损失可用它们的方差比来度量. 于是我们定义 $c'\hat{\beta}$ 关于 $c'\beta^*$ 的相对效率为

$$\text{RE}(c'\hat{\beta}) = \frac{\text{var}(c'\beta^*)}{\text{var}(c'\hat{\beta})}. \tag{9.2.3}$$

它的下界由下面的定理给出. 其证明此处略去了, 感兴趣的读者可以阅读王松桂(1987).

定理 9.2.2　在线性统计模型 (9.1.2) 中, 设 $V>0$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为其特征值. 则

$$\mathrm{RE}(c'\hat{\beta}) \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}. \tag{9.2.4}$$

应用这个定理, 我们容易得到

推论 9.2.1　在定理 9.2.2 条件下

(1) 若 $r(X)=p$, 则

$$\mathrm{cov}(\beta^*) \leqslant \mathrm{cov}(\hat{\beta}) \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} - \mathrm{cov}(\beta^*).$$

(2) 若 $r(X)<p$, 则

$$\mathrm{cov}(X\beta^*) \leqslant \mathrm{cov}(X\hat{\beta}) \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\mathrm{cov}(X\beta^*).$$

在结束这一节的时候, 我们给出不等式 (6.6.6) 在相对效率方面的统计意义, 在线性统计模型 (9.1.2) 中, 设 $V>0$, $r(X)=p$. Rao(1985) 研究了均值向量 $\mu=X\beta$ 的 LS 估计相对于 BLU 估计的效率. 首先注意到, 在研究 $\mu=X\beta$ 的估计时, 不失一般性, 可假设 $X'X=I_p$. 于是 μ 的 LS 估计和 BLU 估计分别为

$$\hat{\mu}=X\hat{\beta}=XX'y,$$

$$\mu^*=X\beta^*=X(X'V^{-1}X)^{-1}X'V^{-1}y.$$

它们的协方差阵分别为

$$\mathrm{cov}(\hat{\mu})=XX'VXX',$$

$$\mathrm{cov}(\mu^*)=X(X'V^{-1}X)^{-1}X'.$$

Rao(1985) 所引进的相对效率的度量为

$$\mathrm{RE}(\hat{\mu})=\mathrm{tr}[\mathrm{cov}(\hat{\mu})-\mathrm{cov}(\mu^*)]. \tag{9.2.5}$$

设 $V=P\Lambda P'$ 为 V 的奇异值分解, 其中 P 为正交阵, $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 为对角阵. 则 (9.2.5) 变形为

$$\mathrm{RE}(\hat{\mu})=\mathrm{tr}[U'\Lambda U-(U'\Lambda^{-1}U)^{-1}],$$

其中 $U = P'X$ 满足 $U'U = I_p$. (6.6.6) 给出了 $\mathrm{Re}(\hat{\mu})$ 的上界

$$\mathrm{RE}(\hat{\mu}) \leqslant \sum_{i=1}^{p}(\lambda_i^{1/2} - \lambda_{n-i+1}^{1/2})^2.$$

Liski, Puntanen 和 Wang(1992) 把上面的结果推广到 $V \geqslant 0$ 的情形.

从不同的角度出发, 我们可以提出 LS 估计相对效率的不同度量. 近年来在这方面出现了不少新工作, 例如王松桂 (1985)、杨虎 (1988)、Wang Songgui 和 Yang Yu(1989)、刘爱义和王松桂 (1991)、王静龙和高道德 (1991) 等.

§9.3 约束的 Kantorovich 不等式及统计应用

本节我们首先把 §4.10 证明过的 Kantorovich 不等式推广到带约束的情形, 然后导出相对效率不等式 (9.2.4) 的一个改进结果及其矩阵形式, 最后介绍它们的一些统计应用.

定理 9.3.1(约束的 Kantorovich 不等式) 设 A 为 $n \times n$ 正定阵, 其特征值为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, $\varphi, \cdots, \varphi_n$ 为对应的标准正交化特征向量, 设 $i_1, \cdots, i_k$ 为正整数, 满足 $1 \leqslant i_1 < \cdots < i_k \leqslant n$, 记 $\boldsymbol{\Phi}_1 = (\varphi_{i_1}, \cdots, \varphi_{i_k})$, 则

$$\sup_{x \in \mathscr{M}(\boldsymbol{\Phi})} \frac{x'Axx'A^{-1}x}{(x'x)^2} = \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}. \tag{9.3.1}$$

证明 不失一般性, 我们假设 $i_e = l$, $l = 1, \cdots, k$, 此时 $\boldsymbol{\Phi}_1 = (\varphi_1, \cdots, \varphi_k)$. 记 $\Lambda_1 = \mathrm{diag}(\lambda_1, \cdots, \lambda_k)$, 则对任意 $x \in \mathscr{M}(\boldsymbol{\Phi}_1)$, 存在 $k \times 1$ 向量 t, 使得 $x = \boldsymbol{\Phi}_1 t$, 于是

$$\sup_{x \in \mathscr{M}(\boldsymbol{\Phi})} \frac{x'Axx'A^{-1}x}{(x'x)^2} = \sup_t \frac{t'\Lambda_1 tt'\Lambda_1^{-1}t}{(t't)^2} = \frac{(\lambda_1 + \lambda_k)^2}{4\lambda_1\lambda_k},$$

这里最后一个等号应用了 Kantorovich 不等式 (4.10.4). 证毕.

注 当 $k = n$ 时, (9.3.1) 就变成无约束的 Kantorovich 不等式 (4.10.4).

现在我们返回到线性统计模型 (9.1.2), 并沿用前面两节的记号, 下面的定理改进了定理 9.2.2.

定理 9.3.2 对线性统计模型 (9.1.2), 设 $r(X) = t$, $V > 0$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 V 的特征值, $\varphi_1, \cdots, \varphi_n$ 为对应的标准正交化特征向量. 设 $\mathscr{M}(X) \subset \mathscr{M}(\varphi_{i_1}, \cdots, \varphi_{i_k})$, 这里 $1 \leqslant i_1 < \cdots < i_k \leqslant n$, 则对一切可估函数 $c'\beta$, 有

$$\mathrm{RE}(c'\hat{\beta}) = \begin{cases} 1, & \text{若 } t = k, \quad (9.3.2) \\ \dfrac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1} + \lambda_{i_k})^2}, & \text{若 } t < k, \quad (9.3.3) \end{cases}$$

这里 $\mathrm{Re}(c'\hat{\beta})$ 的定义同 (9.2.3).

证明 若 $t=k$, 则 $\mathscr{M}(X)=\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$, 应用 LS 估计稳健性理论 (见王松桂 (1987), p. 184, 定理 5.2), 对一切可估函数 $c'\beta$, 它的 LS 估计 $c'\hat{\beta}$ 与 BLU 估计 $c'\beta^*$ 相等, 于是 (9.3.2) 成立.

若 $t<k$, 对 X 作奇异值分解 $X=P\Lambda Q'$, 这里 P 和 Q 分别为 $n\times t$ 和 $p\times t$ 矩阵, 满足 $P'P=I_t$, $Q'Q=I_t$, $\Lambda=\mathrm{diag}(\delta_1,\cdots,\delta_t)$, $\delta_i>0$, $i=1,\cdots,t$, 对每个可估函数 $c'\beta$, 存在 $n\times 1$ 向量 α, 使得 $c=X'\alpha$. 记 $u=P'\alpha$, 利用定理 1.7.5 和 1.7.9, 得

$$\begin{aligned}\mathrm{var}(c'\hat{\beta})&=\alpha'PP'VPP'\alpha=u'PVP'u,\\ \mathrm{var}(c'\beta^*)&=\alpha'X(X'V^{-1}X)^{-1}X'\alpha=u(P'V^{-1}P)^{-1}u\end{aligned}\tag{9.3.4}$$

$$\geqslant\frac{(u'u)^2}{u'P'V^{-1}Pu}.\tag{9.3.5}$$

在最后一步应用了 Cauchy-schwarz 不等式, 由 (9.3.4) 和 (9.3.5) 我们有

$$\begin{aligned}\mathrm{RE}(c'\hat{\beta})&=\frac{(u'u)^2}{u'P'VPuu'P'V^{-1}Pu}=\frac{(\tilde{u}'\tilde{u})^2}{\tilde{u}'V\tilde{u}\tilde{u}'V^{-1}\tilde{u}}\\&\geqslant\frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1}+\lambda_{i_k})^2},\end{aligned}$$

其中 $\tilde{u}=Pu=PP'\alpha\in\mathscr{M}(X)\subset\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$. 上面的不等号是由于定理 9.3.1. 证毕.

因 $\mathscr{M}(X)\subset\mathscr{M}(\varphi,\cdots\varphi_n)$ 总是成立的, 所以 (9.3.3) 包含了 (9.2.4). 在 (9.3.3) 中, 最优下界对应于包含 $\mathscr{M}(X)$ 的最小特征子空间.

上面这两个定理是由本书作者之一和邵军证明的 (见 Wang Songgui 和 Shao Jun(1992)). 该文还列举了许多线性统计模型的例子, 说明它们的设计阵 X 与误差协方差阵的特征向量之间确实存在着形如 $\mathscr{M}(X)\subset\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$ 的关系.

现在我们来导出 Kantorovich 不等式的矩阵形式.

对任一可估函数 $c'\beta$, 存在 α, 使得 $c=X'\alpha$. 如果 $\mathscr{M}(X)\subset\mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$ 成立, 那么由定理 9.3.2, 得

$$\mathrm{RE}(c'\hat{\beta})=\mathrm{RE}(\alpha'X\hat{\beta})\geqslant\frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1}+\lambda_{i_k})^2},$$

即

$$\alpha'P_xVP_x\alpha\leqslant\frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}\alpha'X(X'V^{-1}X)^-X'\alpha,$$

其中 $P_x=X(X'X)^-X'$. 由 α 的任意性, 得

$$P_xVP_x\leqslant\frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}}X(X'V^{-1}X)^-X'.$$

容易验证, 上式等价于

$$X'VX \leqslant \frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} X'X(X'V^{-1}X)^- X'X.$$

用 A^{-1} 代替 V, 我们就证明了如下事实.

定理 9.3.3 设 A 为 $n\times n$ 实对称正定阵. $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值, $\varphi_1,\cdots,\varphi_n$ 为对应的标准正交化特征向量, X 为 $n\times p$ 矩阵, 若存在 $1 \leqslant i_1 < \cdots < i_k \leqslant n$, 使得 $\mathscr{M}(X) \subset \mathscr{M}(\varphi_{i_1},\cdots,\varphi_{i_k})$, 则

$$X'A^{-1}X \leqslant \frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} X'X(X'AX)^- X'X. \tag{9.3.6}$$

特别, 当 $X'X = I_p$ 时,

$$X'A^{-1}X \leqslant \frac{(\lambda_{i_1}+\lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} (X'AX)^{-1}. \tag{9.3.7}$$

因为 $\mathscr{M}(X) \subset \mathscr{M}(\varphi_1,\cdots,\varphi_n)$ 总成立, 所以我们立即得到如下推论.

推论 9.3.1 设 A 为 n 阶实对称正定阵, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为其特征值, X 为任一 $n\times p$ 矩阵, 则

$$X'A^{-1}X \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} X'X(X'AX)^- X'X,$$

特别, 当 $X'X = I_p$ 时,

$$X'A^{-1}X \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} (X'AX)^{-1}.$$

关于 Kantorovich 不等式在这一方向的更一般的推广, 见 Wang Songgui 和 Shao Jun(1992).

§9.4 统计检验

本节将扼要介绍第八章所建立的受控理论在统计检验的容许性、无偏性以及功效函数的单调性等方面的应用.

我们先考虑检验的容许性. 在线性统计模型 (9.1.2) 中, 因变量 Y 只有一个, 如果现在存在 q 个因变量, 设 $q\times 1$ 向量 $y_1,\cdots,y_n$ 为这 q 个因变量的 n 次独立观测, 且每个都服从正态分布 $N(\Theta' x_i, \Sigma)$, $i=1,\cdots,n$. 这里 $p\times 1$ 向量 $x_1,\cdots,x_n$ 为 p 个自变量的 n 组已知值, Σ 和 H 分别为 $p\times p$ 和 $q\times p$ 未知参数阵. 假设 $n > p+q$, 若记

$$\underset{n\times q}{Y} = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix}, \quad \underset{n\times p}{X} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix},$$

则我们有如下多元正态线性统计模型

$$Y = X\Theta + \varepsilon, \tag{9.4.1}$$

这里 ε 为 $n \times q$ 随机误差矩阵, 它的 n 个行相互独立且有相同分布 $N(0, \Sigma)$, 记为 $\varepsilon \sim N(0, \Sigma \otimes I)$. 欲检验的假设为 H: $C\Theta = 0$. 这里 C 为 $r \times p$ 矩阵, $r(C) = r$. 我们的问题是证明常用的几种检验都是容许的, 如果不存在另外的检验比它有更低的水平和更高的功效函数, 这个检验称为是容许的.

模型 (9.4.1) 的典型形式为

$$\begin{aligned}
Z_1 &= M_1 + \varepsilon_1, \quad \varepsilon_1 \sim N(0, \Sigma \otimes I_r);\\
Z_2 &= M_2 + \varepsilon_2, \quad \varepsilon_2 \sim N(0, \Sigma \otimes I_{p-r});\\
Z_3 &= \varepsilon_3, \qquad\quad\ \ \varepsilon_3 \sim N(0, \Sigma \otimes I_{n-p}),
\end{aligned}$$

这里 Z_1, Z_2 和 Z_3 分别为 $r \times q$, $(n-r) \times q$ 和 $(n-p) \times q$ 矩阵, 在典型形式下, 欲检验的假设 H 变为: $M_1 = 0$. 记

$$A = Z_1' Z_1, \quad B = Z_3' Z_3,$$

$f_1 \geqslant \cdots \geqslant f_l$ 为 AB^{-1} 的特征值, $l = \min(q, n-p)$, 常用检验统计量都是 f_i 的函数. 假如

(1) 似然比统计量

$$W = \frac{\det B}{\det(A+B)} = \prod_{i=1}^{l} (1+f_i)^{-1};$$

(2) Lawley-Hotelling 统计量

$$T^2 = \operatorname{tr}(AB^{-1}) = \sum_{i=1}^{l} f_i;$$

(3) Bartlett-Nanda-Pillai 统计量

$$V = \operatorname{tr} A(A+B)^{-1} = \sum_{i=1}^{l} \frac{f_i}{1+f_i};$$

(4) Roy 统计量

$$\lambda_1(AB^{-1}) = f_1.$$

为了证明上述检验都是容许的, 我们需引进单调集合的概念.

记 $D_+ = \{x \in R^n$: $x_1 \geqslant \cdots \geqslant x_n \geqslant 0\}$.

定义 9.4.1 设 $S \subset D_+$, $x \in S$, $y \in D_+$, 若由 $y \prec_w x$, 可推出 $y \in S$, 则称集合 S 是单调的.

引理 9.4.1 设 $S \subset D_+$ 是单调闭凸集, 则 S 是一个容许检验的接受区域.

这个引理的证明从略 (详见 Anderson(1984)).

引理 9.4.2 设 $g(\cdot)$ 为定义在 [0,1] 上的连续非减凸函数, 记

$$\phi(x) = \sum_{i=1}^{n} g(x_i).$$

则以 $S = \{x \in R^n | \phi(x) \leqslant c\}$ 为接受区域的检验是容许的.

证明 由 g 的连续凸性可知, $\phi(x)$ 也是连续的凸函数, 因而 S 为闭凸集, 由推论 8.2.3 知 $\phi(x)$ 为 Schur 凸函数, 于是若 $y \prec_w x$, 有 $\phi(y) \leqslant \phi(x)$. 因而若 $x \in S$, $y \prec_w x$, 必有 $y \in S$. 因而 S 是单调的, 再由引理 9.4.1, 定理得证.

对前面给出的四种检验, 它们的接受区域分别是 $W \leqslant c$, $T^2 \leqslant c$, $V \leqslant c$, $f_1 \leqslant c$. 由引理 9.4.1 可直接验证 Roy 检验是容许的. 对其余检验, 在引理 9.4.2 中, 分别取

$$g(x) = -\log(1+x), \quad g(x) = x, \quad g(x) = \frac{x}{1+x},$$

可知它们都是容许检验. 这就证明了如下定理.

定理 9.4.1 似然比检验、Lawley-Hotelling 检验、Bartlett-Nanda-Pillai 检验和 Roy 检验都是容许的.

上面我们所讨论的是多元正态线性统计模型中的假设检验问题, 现在我们考虑一个一般的统计检验, 假设 X 为一随机变量、向量或矩阵, 其分布函数为 $F(\theta)$, 这里 $\theta = (\theta_1, \cdots, \theta_k)' \in \Omega \subset R^k$. 我们欲检验的统计假设为 H: $\theta \in \Omega_H$. 假定所采用的检验统计量为 $\phi(x)$, 其拒绝区域为 $\{x$: $\phi(x) > c\}$. 该检验的功效函数为

$$\beta(\theta) = P_\theta(\phi(x) > c) = E_\theta I_{(\phi(x)>c)}.$$

设 $0 < \alpha < 1$, 若对一切 $\theta \in \Omega_H$, $\beta(\theta) \leqslant \alpha$, 则称 α 为检验 ϕ 的水平. 进一步, 若对一切 $\theta \overline{\in} \Omega_H$, $\beta(\theta) \geqslant \alpha$, 则称 ϕ 为无偏检验.

对于一个简单情况, H: $\theta_1 = \cdots = \theta_k$, 不难看出检验无偏性的充分条件为

(1) $\beta(\theta)$ 在 Ω 上为 Schur 凸函数, 且当 $\sum_{i=1}^{k} \theta_1$ 固定时, $\beta(\theta)$ 在 $\theta_i = \cdots = \theta_k$ 达到最小;

(2) 假若 $(\theta^*, \cdots, \theta^*) \in \Omega$, $\beta(\theta^*, \cdots, \theta^*)$ 与 θ^* 无关, 于是 $\theta \in \Omega$ 时, $\beta(\theta)$ 为常数.

$\beta(\theta)$ 的 Schur 凸性也保证了功效函数的单调性.

在 (1) 中, 我们要求了 $\beta(\theta)$ 的 Schur 凸性, 事实上, 对于很多分布族, 这一点是满足的. 这是因为当检验统计量 $\phi(x)$ 为 Schur 凸函数时,

$$g(x) = I_{(u,\phi(u)>c)}(x)$$

也是 Schur 凸函数, 对很多分布族 F_θ, $\theta \in \Omega$

$$\beta(\theta) = \int g(x)\mathrm{d}F_\theta(x)$$

也是 Schur 凸函数.

应用上述方法, 可以得到下列检验的无偏性及功效函数的单调性

(1) 多元正态总体的各分量方差相等性检验;

(2) 多元正态总体的协方差阵 $\Sigma = \delta^2 I$ 的检验;

(3) 多元正态总体具有特定协方差结构的检验.

关于这一方向的详细讨论见 Marshall 和 Olkin(1979, p. 386~390).

参考文献

陈希孺, 王松桂. 1987. 近代回归分析 —— 原理、方法及应用 [M]. 安徽教育出版社.

华罗庚. 1955. 一个关于行列式的不等式 [J]. 数学学报 (5): 463~470.

蒋尔雄, 高坤敏, 吴景琨. 1978. 线性代数 [M]. 人民教育出版社.

刘爱义, 王松桂. 1989. 线性模型中最小二乘估计的一种新的相对效率 [J]. 应用概率统计 (5): 97~104.

陆传荣, 林正炎, 陆传赉. 1989. 概率论极限理论引论 [M]. 高等教育出版社.

孙继广. 1987. 矩阵扰动分析 [M]. 科学出版社.

陶波, 成平. 1981. 一个矩不等式 [J]. 数学年刊, 2(4): 451~461.

王松桂. 1985. 广义相关系数与估计效率 [J]. 科学通报 (80): 1521~1524.

王松桂. 1987. 线性模型的理论及其应用 [M]. 安徽教育出版社.

王松桂, 刘爱义. 1989. 两步估计的效率 [J]. 数学学报 (32): 42~54.

王学仁, 王松桂. 1990. 实用多元统计分析 [M]. 上海科技出版社.

王静龙, 高道德等. 1991. 欧氏模意义下广义最小二乘估计的效率 [J]. 应用概率统计 (7): 361~365.

徐利治, 王兴华. 1983. 数学分析的方法及例题选讲 [M]. 高等教育出版社.

许以超. 1965. 代数学引论 [M]. 上海科学技术出版社.

杨虎. 1988. Kantorovich 不等式的延拓与均方误差比效率 [J]. 应用数学 (4): 85~90.

钟开莱. 1989. 概率论教程 [M]. 刘文, 吴让泉, 译. 上海科技出版社.

Anderson T W. 1984. Introduction to Multivariate Statistical Analysis [M]. New York: Wiley.

Baksalary J K. 1986. A relationship between the star and minus ordering [J]. Linear Algebra and its Applications, 82: 163~167.

Baksalary J K, Hauke J. 1984. Inheriting independence and Chi-Squaredness under certain matrix orderings [J]. Statistics and Probability Letter, 2: 35~38.

Baksalary J K, Liski E P, Trenkler G. 1989. Mean square error matrix improvements and admissibility of linear estimator [J]. J. of Statistical Planning and Inference, 23: 313~325.

Baksalary J K, Puntanen S. 1991. Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities. Aequationes Mathematicae [J], 41: 103~110.

Baksalary J K, Schipp B, Trenkler G. 1992. Some further results on Hermitian-matrix inequalities. Linear Algebra and its Applications [J], 160: 119~129.

Baksalary J K, Trenkler G. 1991. Nonnegative and positive definiteness of matrices modified by two matrices of rank one. Linear Algebra and its Applications [J], 151: 169~184.

Beckenbach E F. 1967. A "workshop"on Minkowski's inequality. In "Inequalities"[C]. ed., O.Shisha. New York: Academic Press, 37~55.

Bellman R. 1970. Introduction to Matrix Analysis [M]. Mcgraw-Hill Book Company.

Bennett G. 1965. Upper bounds on the moments and probability inequalities for the sum of independent [J]. Biometrka: bounded random variables, 52: 559~569.

Besly A, Kuh E, Welsch R E. 1980. Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity [M]. New York: John Wiley.

Bhatia R. 1997. Matrix Analysis [M]. New York: Springer-Verlag.

Bloomfield P, Watson G S. 1975. The inefficiency of least squares [J]. Biometrika, 62: 121~128.

Bushell P J, Trustrum G B. 1990. Trace inequalities for positive definite matrix power products [J]. Linear Algebra and its Applications, 132: 173~178.

Chan N N, Kwong M K. 1985. Hermitian matrix inequalities and a conjecture [J]. American Mathematical Monthly, 2: 533~541.

Courant R. 1992. Zur Theorie der kleinen schwingungen. Z [J]. Angew. Math. Mech., 2: 278~285.

Diaz J B, Metcalf F T. 1964. Complementary inequalities for sums of real numbers [J]. J. of Mathematical Analysis and Applications, 9: 59~74.

Diaz J B, Metcalf F T. 1967. Inequalities complementary to Cauchy's inequality for sums of real number. In "Inequalities" [C], ed., O. Shisha. New York: Academic Press, 73~77.

Dixon J D. 1984. How good is Hadamard's inequality for determinants [J]. Canad. Math. Bull., 27: 260~264.

Drazin M P. 1978. Natural structures on semigroups with involution [J]. Bulletin of American Mathematical Society, 84: 139~141.

Durbin J, Watson G S. 1950. Testing for serial correlation in least squares regression. I. Biometrika, 37: 409~428. Corrections, 1951, 38: 177~178.

Fan K, Hoffman A. 1955. Some metric inequalities in the space of matrices [J]. Proc. Am. Math. Soc. 6: 111~116.

Fisz M. 1963. Probability Theory and Mathematical Statistics [M]. New York: John Wiley.

Gnedenko B V. 1962. The Theory of Probability [M]. New York: Chelsea Pub. Co..

Greub W, Rheinbpldt W. 1959. On a generalization of an inequality of L. V. Kantorovich [J]. Proc. Amer. Math. Sos., 10: 407~415.

Gurland J. 1967. An inequality Satisied by the expectation of the reciprocal of a random variable [J]. Amer. Statist., 21(2): 24~25.

Gurland J. 1968. Inequalities of expectations of random variables derived by monotonility or convexity [J]. Ame. Statist., 22(2): 26~27.

Hardy G H, Littlewood J E, Polya G. 1934, 1952. Inequalities. 1st ed; 2nd ed. London: Cambridge Univ. Press.

Hartwig R E. 1980. How to partially order regular elements [J]. Math. Japonica, 25: 1~13.

Hartwig R E, Styan G P H. 1987. Partially ordered idempotent matrices. in Proceedings of the Second International Tampere Conference in Statistics [C]. ed Pukkila T and Puntanen S., Tampere: Univ. of Tampere, 361~383.

Haynesworth E V. 1957. Note on bounds for certain determinants [J]. Duke Math. J., 24: 313~320.

Haynesworth E V. 1960. Bounds for Determinants with positive diagonals [J]. Trans. Amer. Math. Soc., 96: 395~413.

Hogg R V, Craig A T. 1970. Introduction to Mathematical Statistics [M], New York: The Mac Millan Co.

Karhn S, Studen W J. 1966. Tchebyohef's System with Applications in Analysis and Statistics [M]. New York: Interscience Publishers.

Keilson J. 1972. A threshold for Iog-covarity for probability generating functions and associated moment inequalities [J]. Ann. Math. Statist., 43: 1702~1708.

Kendall M G, Stuart A. 1958. The Advanced Theory of Statistics I. New York: Hafrer Pub. co.

Khatri C G. 1983. A generalization of Lavoie's inequality concerning the sum of idempotent matrices [J]. Linear Algebra and its Applications, 54: 97~108.

Khatri C G, Rao C R. 1981. Some extensions of the Kantorovich inequality and statistical applications [J]. J. Multi. Annalysis, 11: 498~505.

Kimeldorf G, Sampson A. 1973. A class of covariance inequalities [J]. J. Amer. Statist. Asso., 68: 228~230.

Klamkin M S. 1975. Extension of the Weierstrass product inequalities [J]. II. Amer. Math. Monthly 82: 741~742.

Klamkin M S, Newman D J. 1970. Extension of the Weierstrass product inequalities [J]. Math. Mag, 43: 137~141.

Kounias E G, Weng T S. 1969. An equality and almost sure convergence [J]. Ann. Math. Statist., 40: 1091~1093.

Lavoie J L. 1980. A determinantal inequality involving the Moore-Penrose inverse [J]. Linear Algebra and its Applications, 31: 78~80.

Lei T G, Woo C W, Zhang Fu-Zhen. 2004. Matrix inequalities by means of embedding. Electronic Journal of Linear Algebra. ISSN 1081~3810. A publication of the International Linear Algebra Society, 11: 66~77.

Lidskii V. 1950. The proper values of the sum and product of symmetric matrices [J]. Dokl. Akad. Nauk SSSR, 75: 769~772.

Liski E P. 1990. On Lowner-ordering antitonicity of matrix inversion [J]. Technical Report. A: 427. Dept. of Math. Univ. of Tampere.

Liski E P, Puntanen S, Wang S G (王松桂). 1992. Bounds for the trace of the difference of the covariance of the OLSE and BLUE[J]. Linear Algebra and its Applications, 176: 121~130.

Liski E P, Wang Song-Gui(王松桂). 1991. Effects of observations on the eigensystem of a sample covariance matrix [J]. J. of Statistical Planning and Inference, 36: 215~226.

Liu Shuangzhe and Neudecker Heinz. 1996. Several Matrix Kantorovich-Type Inequalities [J]. Journal of Mathematical Analysis and Application, 197: 23~26.

Loeve M. 1960. Probability Theory [M]. New Jersey. Priuceton: D. Van Nostrand Co. Inc.

Magnus J R. 1987. A representation theorem for (tr $A^p)^{1/p}$[J]. Linear Algebra and its Applications, 95: 127~134.

Malamud S M. 2001. A converse to the Jensen inequality, its matrix extensions and inequalities for minors and eigenvalues [J]. Linear Algebra and its Applications, 322: 19~41.

Mallows C L, Richter D. 1969. Inequalities of Chebyshev type involving conditional expectation [J]. Ann. Math. Statist., 40: 1922~1932.

Marcus M. 1956. An eigenevalue inequality for the product of normal matrices [J]. Amer. Math. Monthly, 63: 173~174.

Marsaglia G, Styan G P H. 1974a. Rank conditions for generalized inverse of partitioned matrices [J]. Sankhya. Ser., A 36: 437~442.

Marsaglia G, Styan G P H. 1974b. Equalities and inequalities for ranks of matrices [J]. Linear and Multilinear Algebra, 2: 269~292.

Marshall A W, Olkin I. 1965. Norms and inequalities for condition numbers [J]. Pacific J. of Mathematics, 15: 241~247.

Marshall A W, Olkin I. 1979. Inequalities: Theory of Majorization and its Applications [M]. New York: Academic Press.

Marshall A W, Olkin I. 1990. Matrix versions of the Cauchy and Kantorovich inequalities [J]. Aequaliones Mathematiae, 40: 89~93.

McCarthy Ch, Stang C. 1973. Optimal conditioning of matrices [J]. SIAM. J. Numer. Anal., 10: 370~388.

Mirsky L. 1955. An Introduction to Linear Algebra [M]. Oxford University Press.

Mitra S K. 1986. The minus partial order and the shorted matrix [J]. Linear Algebra and its Applications, 83: 1~27.

Mitrinovic D S. 1970. Analytic Inequalities [M]. New York: Springerverlag Heidelbery.

Mond B, Pečarić J E. 1995a. Generalization of a matrix inequality of Ky Fan [J]. Journal of Mathematical Analysis and Application, 190: 244~247.

Mond B, Pečarić J E. 1995b. A matrix version of the Ky Fan generalization of the Kantorovich inequality [J]. II. Linear and Multilinear Algebra, 38: 309~313.

Mond B, Pečarić J E. 1995c. Reverse Forms of a convex matrix inequality [J]. Linear Algebra and its Applications, 220: 359~364.

Mond B, Pečarić J E. 1997. Matrix inequalities for convex functions [J]. Journal of Mathematical Analysis and Application, 209: 147~153.

Mond B, Shisha O. 1967. Ratios of means and applications. In "inequalities" [C], ed., Oshi. sha, New York: Academic Press.

Moors J J A, Muilwijh. 1971. An inequality of variance of a discrete random variable [J]. Sankhya B, 33: 385~388.

Muir T. 1906. The Theory of Determinants in the Historical Order of Development [M]. New York: MacMillan.

Mullen K. 1967. A note on the ratio independent random variables [J]. Amer. Statist, 21(3): 30~31.

Patel J K, Kapadia C H, Owen D B. 1976. Handbook of Statistical Distribution [M]. New York: Marcel Dekker, Inc.

Poincare H. 1890. Sur les equations aux derivees partielles de la physique mathematique [J], Amer. Math., 12: 211~294.

Rao C R. 1973. Linear Statistical Inferences. New York: Wiley.

Rao C R. 1985. The inefficiency of least square: extensions of the Kantorovich inequality [J]. Linear Algebra and its Applications. 70: 209~224.

Savage I R. 1961. Probabilty inequalities of the Tchebycheff tpye [J]. J. Res. Nat. Bur. Stand., 65(B): 211~222.

Scott A J, Styan G P H. 1985. On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys [J]. Linear Algebra and its Applications, 70: 209~224.

Serffing R J. 1980. Approximation Theorems of Mathematical Statistics [M]. New York: Wiley.

Shapiro A. 1982. Optimally scaled matrices, necessary and sufficient conditions [J]. Numer. Math., 39: 239~245.

Stepniak C. 1985. Ordering of nonnegative definite matrices with application to comparison of linear models [J]. Linear Algebra and its Applications, 70: 67~71.

Stepniak C, Wang S G (王松桂), Wu C F. 1984. Comparison of linear experiments with known variance [J]. Ann. of Statistics, 12: 358~365.

Styan G P H. 1981. On Lavoiés determinantal inequality [J]. Linear Algebra and its Applications, 37: 77~80.

Tomkins S J. 1971. On an equality of Heyde [J]. J. Appl. Prob. 8: 428~429.

Tong Y L. 1970. Some probability inequalities of multivariate normal and multivariate [J]. J. Amer. Statist. Asso., 65: 1243~1247.

Tukey J W. 1946. An equality for deviatials from medians [J]. Ann. Math. Statist, 17: 75~78.

Wang Bo-Ying, Xi Bo-Yan, Zhang Fuzhen. 1999. Some inequalities for sum and product of positive semidefinite matrices. Linear and Multilinear Algebra, 293: 39~49.

Wang Boying(王伯英), Zhang Fuzhen(张福振). 1992. Some inequalities for the eigenvalues of the product of positive semi-definite Hermitian Matrices [J]. Linear Algebra and its Applications, 160: 113~118.

Wang Song-Gui(王松桂), Chow Shein-Chung. 1993. Advanced Linear Models: Theory and Applications [M]. New York: Marcel Dekker Inc.

Wang Song-Gui(王松桂), Sin-keung Tse and Shein-Chung Chow. 1990. On the measures of multicollinearity in least squares regression [J]. Statistics and Probability Letter, 9: 347~355.

Wang Song-Gui(王松桂), Shao J.(邵军). 1992. Constrained Kantorovich Inequalities and relative efficiency of least squares [J]. J. Multi. Analysis, 42: 284~298.

Wang Song-Gui(王松桂), Yang Hu(杨虎). 1989. Kantoroyicg-tpye inequalities and the measures of inefficiency of the GLSE [J]. Acta Mathematicae Applicatae Sinica, 5: 372~381.

Wang Song-Gui(王松桂), Ip Wai-Cheung 1999. A matrix version of the Wielandt inequality and its applications to statistics [J]. Linear Algebra and its Applications, 296: 171~181.

Wu Chien-Fu. 1980. On some ordering properties of the generalized inverses of nonnegative definite matrices [J]. Linear Algebra and its Applications, 32: 49~60.

Yang X J. 2000. NOTE: a matrix trace inequality [J]. Journal of Mathematical Analysis and Applications, 250: 372~374.

Zhan X Z. 2000. Singular values of differences of positive semidefinite matrices [J]. SIAM J. Matrix Analysis and Appl., 22(3): 819~823.

Zhang Fuzhen. 2001. Matrix inequalities by means of block matrices [J]. Mathematical Inequalties and Applications, 4: 481~490.

附录 1　关于数量和函数的不等式

1. 平均值不等式

设 $a_1, \cdots, a_n$ 为非负实数, 对 $r \neq 0$, 定义

r 次幂平均: $M_r(a) = \left(\frac{1}{n}\sum_{i=1}^{n} a_i^r\right)^{1/r}$,

$$M_r(a) = 0, \quad \text{当 } r < 0 \text{ 且至少一个 } a_i = 0;$$

算术平均: $A(a) = M_1(a) = \frac{1}{n}\sum_{i=1}^{n} a_i$;

调和平均: $H(a) = M_{-1}(a) = \frac{n}{\sum_i \frac{1}{a_i}}$;

几何平均: $G(a) = \sqrt[n]{a_1 \cdots a_n}$.

(1) 调和平均 - 几何平均 - 算术平均不等式

$$H(a) \leqslant G(a) \leqslant A(a);$$

(2) 设 $r < s$, 则 $M_r(a) \leqslant M_s(a)$;

(3) 当 $n = 2$ 时, $\frac{1}{2}(A(a) + H(a)) \geqslant G(a)$;

(4) 当 $a_1, \cdots, a_n$ 为不全相等的正数时,

$$1 < \frac{A(a) - H(a)}{A(a) - G(a)} < n;$$

(5) 设 $q_i \geqslant 0$, $i = 1, \cdots, n$, $\sum_{i=1}^{n} q_i = 1$, 则

$$\sum_{i=1}^{n} a_i q_i \geqslant \prod_{i=1}^{n} a_i^{q_i}. \qquad \text{(徐利治, 王兴华　1983)}$$

2. Abel 不等式

设 $a, b \in R^n$, $b_1 \geqslant \cdots \geqslant b_n \geqslant 0$, 定义 $S_k = \sum_{i=1}^{k} a_i$, $k = 1, \cdots, n$, $m = \min_k S_k$, $M = \max_k S_k$, 则

$$mb_1 \leqslant a'b \leqslant Mb_1. \qquad \text{(Mitrinovic　1970, p. 32)}$$

3. Beckenbach 不等式

设 $1 \leqslant p \leqslant 2$, a_i, $b_i > 0$, $i = 1, \cdots, n$, 则

$$\frac{\sum_{i=1}^{n}(a_i + b_i)^p}{\sum_{i=1}^{n}(a_i + b_i)^{p-1}} \leqslant \frac{\sum_{i=1}^{n} a_i^p}{\sum_{i=1}^{n} a_i^{p-1}} + \frac{\sum_{i=1}^{n} b_i^p}{\sum_{i=1}^{n} b_i^{p-1}}.$$

(Beckenbach, Bellman 1983, p. 27)

4. Bernoulli 不等式

(1) 设 $x > -1$, n 为正整数, 则

$$(1 + x)^n \geqslant 1 + nx;$$

(2) $-2 \leqslant x \leqslant -1$, n 为正整数, 则

$$(1 + x)^n \geqslant 1 + x \geqslant 1 + nx;$$

(3) 若 $n = 2, 3, \cdots, -1 < x < \dfrac{1}{n-1}$, 则

$$(1 + x)^n \leqslant 1 + \frac{nx}{1 + (1 - n)x},$$

等号成立$\Longleftrightarrow x = 0$;

(4) $x > -1$, $0 < \alpha < 1$, 则

$$(1 + x)^\alpha \leqslant 1 + \alpha x;$$

(5) $x > -1$, $\alpha < 0$ 或 $\alpha > 1$, 则

$$(1 + x)^\alpha \geqslant 1 + \alpha x.$$

(前三条见 Mitrinovic 1970, p. 34~35, 后两条见徐利治等 1983, p. 171)

5. Carleman 不等式

设 $a_i \geqslant 0$, $i = 1, 2, \cdots$, 且 $\sum_{i=1}^{\infty} a_i < \infty$, 则

$$\sum_{i=1}^{\infty}(a_1, \cdots, a_n)^{1/n} \leqslant e \sum_{i=1}^{\infty} a_i.$$

(徐利治, 王兴华 1983, p. 158)

6. Cassels 不等式

设 $a_i, b_i > 0$, $w_i \geqslant 0$, $i = 1, \cdots, n$, 且 w_i 不全为零, 则

$$1 \leqslant \frac{\sum_{i=1}^{n} w_i a_i^2 \cdot \sum_{i=1}^{n} w_i b_i^2}{\left(\sum_{i=1}^{n} w_i a_i b_i\right)^2} \leqslant \max_{i,j} \frac{(a_i b_j + a_j b_i)^2}{4 a_i a_j b_i b_j}.$$

(Beckenbach, Bellman 1983, p. 45)

7. Chebychev 不等式

设 $a_1 \leqslant \cdots \leqslant a_n$, $b_1 \leqslant \cdots \leqslant b_n$ 或 $a_1 \geqslant \cdots \geqslant a_n$, $b_1 \geqslant \cdots \geqslant b_n$, 则

$$\left(\frac{1}{n} \sum_{i=1}^{n} a_i\right) \left(\frac{1}{n} \sum_{i=1}^{n} a_i\right) \leqslant \frac{1}{n} \sum_{i=1}^{n} a_i b,$$

等号成立 $\Longleftrightarrow a_1 = \cdots = a_n$, $b_1 = \cdots = b_n$. (徐利治, 王兴华 1983, p. 126)

8. Diaz-Metcalf 不等式

(1) 设 $a_i \neq 0$, b_i $(i = 1, \cdots, n)$ 均为实数, 且

$$m \leqslant \frac{b_i}{a_i} \leqslant M, \quad i = 1, \cdots, n.$$

则

$$\sum_{i=1}^{n} b_i^2 + mM \sum_{i=1}^{n} a_i^2 \leqslant (m + M) \sum_{i=1}^{n} a_i b_i \leqslant |M + m| \left(\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2\right)^{1/2},$$

等号成立 $\Longleftrightarrow b = m a_i$ 或 $b_i = M a_i$, $i = 1, \cdots, n$. (Diaz-Metcalf 1964, p. 62)

(2) 设 $0 \leqslant m_1 \leqslant a_i \leqslant M_1$, $0 \leqslant m_2 \leqslant b_i \leqslant M_2$, x_i 为实数, 则

$$\begin{aligned} m_1 M_1 \sum_{i=1}^{n} b_i^2 x_i^2 + m_2 M_2 \sum_{i=1}^{n} a_i^2 x_i^2 &\leqslant (M_1 M_2 + m_1 m_2) \sum_{i=1}^{n} a_i b_i x_i^2 \\ &\leqslant (M_1 M_2 + m_1 m_2) \left(\sum_{i=1}^{n} a_i^2 x_i^2 + \sum_{i=1}^{n} b_i^2 x_i^2\right)^{1/2}, \end{aligned}$$

当 $m_1 > 0$, $m_2 > 0$ 时, 左边等号成立 $\Longleftrightarrow$ 对每个 $x_i \neq 0$ 的 i, 有 $(a_i, b_i) = (m_1, M_2)$ 或 $(a_i, b_i) = (M_1, m_2)$ 之一成立. (Diaz, Metcalf 1964, p. 62)

9. Dresher 不等式

设 $p \geqslant 1 \geqslant r \geqslant 0$, $f \geqslant 0$, $g \geqslant 0$, 则

$$\left(\frac{\int |f+g|^p dx}{\int |f+g|^r dx}\right)^{1/(p-r)} \leqslant \left(\frac{\int f^p dx}{\int f^r dx}\right)^{1/(p-r)} + \left(\frac{\int g^p dx}{\int g^r dx}\right)^{1/(p-r)}.$$

(Beckenbach, Bellman 1983, p. 28)

10. Fan Ky 不等式

设 $0 < a_i \leqslant \frac{1}{2}$, $i = 1, \cdots, n$, 则

$$\frac{\prod_{i=1}^{n} a_i}{\left(\sum_{i=1}^{n} a_i\right)^n} \leqslant \frac{\prod_{i=1}^{n}(1-a_i)}{\left(\sum_{i=1}^{n}(1-a_i)\right)^n},$$

等号成立 $\Longleftrightarrow a_1 = \cdots = a_n$. (徐利治, 王兴华 1983, p. 162)

11. Fan Ky-Todd 不等式

设 a_i, b_i, $i = 1, \cdots, n$ 皆为实数, $a_j b_i \neq a_i b_j$, $i \neq j$, 则

$$\frac{\binom{n}{2}^2 \sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} a_i b_i\right)^2} \leqslant \sum_{i=1}^{n}\left(\sum_{j \neq i} \frac{a_j}{a_j b_i - a_i b_j}\right)^2.$$

(Beckenbach, Bellman 1983, p. 45)

12. Greub-Rheinboldt 不等式

设 $0 < m_1 \leqslant a_i \leqslant M_1$, $0 < m_2 \leqslant b_i \leqslant M_2$, $x_i (i = 1, \cdots, n)$ 为实数, 则

$$\sum_{i=1}^{n} a_i^2 x_i^2 \sum_{i=1}^{n} b_i^2 x_i^2 \leqslant \frac{(M_1 M_2 + m_1 m_2)^2}{4 m_1 m_2 M_1 M_2}\left(\sum_{i=1}^{n} a_i b_i x_i^2\right)^2.$$

(Grub, Rheinboldt 1959)

13. Hardy 不等式

设 $a_i \geqslant 0$, $i = 1, 2, \cdots$, $S_k = \sum_{i=1}^{k} a_i$, $k = 1, 2, \cdots$.

对 $p>1$, 有

$$\sum_{i=1}^{\infty}\left(\frac{S_k}{k}\right)^p \leqslant \left(\frac{p}{p-1}\right)^p \sum_{k=1}^{\infty} a_k^p.$$

(徐利治, 王兴华　1983, p. 157)

14. Holder 不等式

(1) 设 $a_i \geqslant 0, b_i \geqslant 0, i=1,\cdots,n, p>1, \dfrac{1}{p}+\dfrac{1}{q}=1$. 则

$$\sum_{i=1}^{n} a_i b_i \leqslant \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} \cdot \left(\sum_{i=1}^{n} b_i^q\right)^{1/q},$$

等号成立 $\Longleftrightarrow$ 存在数 c, 使得 $a_i^p = c b_i^q, i=1,\cdots,n$,

等价地, 当 $0<\alpha<1$ 时,

$$\sum_{i=1}^{n} a_i^{\alpha} b_i^{1-\alpha} \leqslant \left(\sum_{i=1}^{n} a_i\right)^{\alpha} \left(\sum_{i=1}^{n} b_i\right)^{1-\alpha},$$

等号成立 $\Longleftrightarrow$ 存在数 c, 使得 $a_i = c b_i, i=1,\cdots,n$.

当 $a_i>0, b_i>0, i=1,\cdots,n, p<1, \dfrac{1}{p}+\dfrac{1}{q}=1$ 时, 不等号反向.

(2) 设 $a_{ij} \geqslant 0, i=1,\cdots,n; j=1,\cdots,m, a_1,\cdots,a_n$ 皆为正数, 且 $a_1+\cdots+a_n=1$.

$$\sum_{j=1}^{m} a_{1j}a_{2j}\cdots a_{nj} \leqslant \left(\sum_{j=1}^{m} a_{1j}^{\frac{1}{a_1}}\right)^{a_1} \cdot \left(\sum_{j=1}^{m} a_{2j}^{\frac{1}{a_2}}\right)^{a_2} \cdots \left(\sum_{j=1}^{m} a_{nj}^{\frac{1}{2n}}\right)^{a_n},$$

等号成立 $\Longleftrightarrow a_1,\cdots,a_n$ 线性相关, 这里 $a=(a_{i1},\cdots,a_{im})'$, $i=1,\cdots,n$.

(Beckenbach, Bellman　1983, p19~20)

(3) 积分形式. 设 $p>1, \dfrac{1}{p}+\dfrac{1}{q}=1$, 则

$$\int fg dx \leqslant \left(\int |f|^p dx\right)^{1/p} \left(\int |g|^q dx\right)^{1/q}.$$

(Beckenbach, Bellman　1983, p. 21)

15. Jesen 不等式

设 $a_1,\cdots,a_n$ 皆为正数, $0<r<p$, 则

$$\left(\sum_{i=1}^{n} a_i^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} a_i^r\right)^{1/r}.$$

(徐利治, 王兴华　1983, p. 127)

16. Just-Schaumberger 不等式

设 $x_i > 0,\ i = 1, \cdots, n,\ s = \sum\limits_{i=1}^{n} x_i$, 则

$$\sum_{i=1}^{n} \frac{s - x_i}{x_i} \geqslant n(n-1).$$

(Marshall, Olkin　1979, p. 72)

17. Kantorovich 不等式

设 $0 < m \leqslant \lambda_i \leqslant M,\ i = 1, \cdots, n,\ x_i$ 皆为实数, 则

$$\left(\sum_{i=1}^{n} \lambda_i x_i^2\right)\left(\sum_{1=i}^{n} \lambda_i^{-1} x_i^2\right) \leqslant \frac{(M+m)^2}{4mM}\left(\sum_{i=1}^{n} x_i^2\right)^2.$$

(Diaz, Metcalf　1967, p. 74)

18. Klamkin-Newman 不等式

设 $x_i > 0,\ i = 1, \cdots, n,\ \sum\limits_{i=1}^{n} x_i = 1$, 则

$$\prod_{i=1}^{n} \frac{1 + x_i}{x_i} \geqslant (n+1)^n,$$

$$\prod_{i=1}^{n} \frac{1 - x_i}{x_i} \geqslant (n-1)^n.$$

(本书例 8.2.6)

19. Klamkin 不等式

设 $x_i > 0,\ i = 1, \cdots, n,\ \sum\limits_{i=1}^{n} x_i = 1$, 则

$$\prod_{i=1}^{n} \frac{1 + x_i}{1 - x_i} \geqslant \left(\frac{n+1}{n-1}\right)^n.$$

(本书例 8.2.6)

20. Laplace 不等式

设 $0 < x_1 < \cdots < x_n,\ 0 < y_1 < \cdots < y_n$, 则

$$\frac{\sum\limits_{i=1}^{n} x_i^2 y_i^2}{\sum\limits_{i=1}^{n} x_i y_i} > \frac{\sum\limits_{i=1}^{n} x_i^2}{\sum\limits_{i=1}^{n} x_i}.$$

(徐利治, 王兴华　1983, p. 141)

21. Minkowski 不等式

(1) 设 $a_i \geqslant 0$, $b_i \geqslant 0$, $i = 1, \cdots, n$, $p > 1$, 则

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n}a_i^p\right)^{1/p} + \left(\sum_{i=1}^{n}b_i^p\right)^{1/p},$$

等号成立 $\iff$ 存在数 c, 使得 $a_i = cb_i$, $i = 1, \cdots, n$. 当 $0 < p < 1$ 不等号反向.

(Beckenbach, Bellman　1983, p. 25~26)

(2) 设 $a_i \geqslant 0$, $b_i \geqslant 0$, $i = 1, \cdots, n$.

$$\left(\prod_{i=1}^{n}(a_i+b_i)\right)^{1/n} \geqslant \left(\prod_{i=1}^{n}a_i\right)^{1/n} + \left(\prod_{i=1}^{n}b_i\right)^{1/n}.$$

(同上, p. 26)

(3) 设 $a_{ij} \geqslant 0$, $i = 1, \cdots, m$; $j = 1, \cdots, n$, $p > 1$.

$$\left(\sum_{j=1}^{n}(a_{1j}+\cdots+a_{mj})^p\right)^{1/p} \leqslant \left(\sum_{j=1}^{n}a_{j1}^p\right)^{1/p} + \cdots + \left(\sum_{j=1}^{n}a_{mj}^p\right)^{1/p}.$$

当 $p < 1$ 时, 不等号反向 (对 $p < 0$, 需假设 $a_{ij} > 0$). 在任何一种情况下, 等号成立 $\iff a_1, \cdots, a_m$ 线性相关, 此处 $a_i = (a_{i1}, \cdots, a_{in})'$, $i = 1, \cdots, m$.　　(同上, p. 21)

(4) 积分形式. 设 $p > 1$.

$$\left(\int |f+g|^p dx\right)^{1/p} \leqslant \left(\int |f|^p dx\right)^{1/p} + \left(\int |g|^p dx\right)^{1/p}.$$

(同上, p. 21)

(5) 推广形式. 设 $c_1, \cdots, c_m$, $b_1, \cdots, b_n$ 为给定的正数, $0 < m < n$, 则对 $p > 1$ 及一切正数 $x_{m+1}, \cdots, x_n$, 下面的不等式成立.

$$\left(\sum_{i=1}^{m}c_i^p + \sum_{i=m+1}^{n}x_i^p\right)^{1/p} - \left[\sum_{i=1}^{m}(c_i+b_i)^p + \sum_{i=m+1}^{n}(x_i+b_i)^p\right]^{1/p}$$

$$\geqslant \left(\sum_{i=1}^{m}c_i^p + \sum_{i=m+1}^{n}d_i^p\right)^{1/p} - \left[\sum_{i=1}^{m}(c_i+b_i)^p + \sum_{i=m+1}^{n}(d_i+b_i)^p\right]^{1/p},$$

其中 $d_i = c^* b_i,\ i = m+1,\cdots,n.$

$$c^* = \frac{\left(\sum_{i=1}^{m} c_i^p\right)^{1/p}}{\left(\sum_{i=1}^{m}(c_i+b_i)^p\right)^{1/p} - \left(\sum_{i=1}^{m} c_i^p\right)^{1/p}}.$$

当 $p<1$ 且 $p\neq 0$ 时, 不等号反向. 对所有情况, 等号成立 $\Longleftrightarrow x_i = d_i,\ i = m+1,\cdots,n.$

注 1　当 $m=1$, 对给定的正数 c_1 和 $b_1,\cdots,b_n$, $p>1$ 及一切正数 $x_2,\cdots,x_n$, 有

$$\left(c_1^p + \sum_{i=2}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} b_i^p\right)^{1/p} \geqslant \left[(c_1+b_1)^p + \sum_{i=2}^{n}(x_i+b_i)^p\right]^{1/p}.$$

当 $p<1$, $p\neq 0$ 时, 不等号反向. 对所有情况, 等号成立当且仅当

$$x_i = \frac{c_1}{b_1} b_i,\quad i = 2,\cdots,n.$$

(Beckenbach　1967, p. 44)

22. 设 $a_1,\cdots,a_n$, $b_1,\cdots,b_n$ 为正数, $p>1$, $\frac{1}{p}+\frac{1}{q}=1$, $0<m<M$,

$$m \leqslant \frac{a_i^{1/q}}{b_i^{1/p}} \leqslant M,\qquad i = 1,\cdots,n.$$

记 $\theta = M/m$, 则

$$1 \leqslant \frac{\left(\sum_{i=1}^{n} a_i^p\right)^{1/p}\left(\sum_{i=1}^{n} b_i^q\right)^{1/q}}{\sum_{i=1}^{n} a_i b_i} \leqslant \left[\frac{q}{p+q}\,\frac{\theta^p-\theta^{-q}}{1-\theta^{-q}}\right]^{1/p}\left[\frac{p}{p+q}\,\frac{\theta^p-\theta^{-q}}{\theta^p-1}\right]^{1/q},$$

等号成立当且仅当所有的 $a_i^{1/q}/b_i^{1/p}$ 都相等.

(Mond, Shisha　1967, p. 193)

23. 设 $a_1,\cdots,a_n$, $b_1,\cdots,b_n$ 为正数, $p>1$, $\frac{1}{p}+\frac{1}{q}=1$, $0<m<M$, 则

$$m \leqslant \left(\frac{a_i}{a_i+b_i}\right)^{1/q} \leqslant M,$$

$$m \leqslant \left(\frac{b_i}{a_i+b_i}\right)^{1/q} \leqslant M,$$

记 $\theta = \dfrac{M}{m}$, 则

$$1 \leqslant \frac{\left(\sum_{i=1}^{n} a_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} b_i^p\right)^{1/p}}{\left(\sum_{i=1}^{n} (a_i+b_i)^p\right)^{1/p}} \leqslant \left[\frac{q}{p+q}\frac{\theta^p-\theta^{-q}}{1-\theta^{-q}}\right]^{1/p}\left[\frac{p}{p+q}\frac{\theta^p-\theta^{-q}}{\theta^p-1}\right]^{1/q}.$$

(Mond, Shisha　1967, p. 194)

24. Schur 不等式

设 $a_{ij} \geqslant 0$, $i,j=1,\cdots,n$, $\sum_{i=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} = 1$, 又 $x_i \geqslant 0$, $i=1,\cdots,n$. 定义 $y_j = \sum_{j=1}^{n} a_{ij}x_j$, 则

$$\prod_{i=1}^{n} y_i \geqslant \prod_{i=1}^{n} x_i.$$

(徐利治, 王兴华　1983, p. 156)

注 2　这个不等式的另一种表述: 设 $x,y \in R^n$, $x_i, y_i \geqslant 0$, $y < x$, 则 $\prod_{i=1}^{n} y_i \geqslant \prod_{i=1}^{n} x_i$. 见本书例 8.2.7.

25. Polya-Szego 不等式

设 $0 < m_1 \leqslant a_i \leqslant M_1$, $0 < m_2 \leqslant b_i \leqslant M_2$, $i=1,\cdots,n$, 则

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \leqslant \frac{(M_1M_2+m_1m_2)^2}{4m_1m_2M_1M_2}\left(\sum_{i=1}^{n} a_ib_i\right)^2.$$

注 3　这是 Greub-Rheinboldt 不等式 $x_i = 1$, $i=1,\cdots,n$ 的特例.

(同上, p. 21)

26. Young 不等式

设 $\phi(x)$ 为 $x(\geqslant 0)$ 的严格单调增连续函数, $\phi(0)=0$, $a \geqslant 0$, $b \geqslant 0$, 则

$$ab \leqslant \int_0^a \phi(x)\mathrm{d}x + \int_0^b \phi^{-1}(x)\mathrm{d}x.$$

等号成立 $\Longleftrightarrow b = \phi(a)$, 其中 $\phi^{-1}(x)$ 表示 $\phi(x)$ 的反函数.

特别, $\phi(x) = x^{p-1}$, $p > 1$, 得

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}, \qquad q = \frac{p}{p-1}.$$

当 $\phi(x) = \ln(x+1)$ 时,

$$ab \leqslant a\ln a - a + e^b.$$

(Beckenbach, Bellman　1983, p. 15)

27. 设 $x_i > 0, i = 1, \cdots, n$, 则

$$\sum_{i=1}^{n} \left(x_i + \frac{1}{x_i}\right)^a \geqslant \frac{(n^2+1)^a}{n^{a-1}}.$$

(本书例 8.2.4)

28. 设 $x_i > 0, i = 1, \cdots, n, \sum_{i=1}^{n} x_i = 1$, 则

$$\sum_{i \neq j} \frac{x_i x_j}{x_i + x_j} \leqslant \frac{n-1}{2}.$$

(Marshall, Olkin　1979, p. 91)

29. 设 $x_i \geqslant 0, i = 1, \cdots, n, M(x) = \frac{1}{n} \sum_{i=1}^{n} x_i, G(x) = \sqrt[n]{x_1 \cdots x_n}$, 则

$$\sum_{i=1}^{n} \log(1 + x_i) \geqslant n \log(1 + M(x)) \geqslant n \log(1 + G(x)).$$

即

$$\prod_{i=1}^{n} (1 + x_i) \geqslant (1 + G(x))^n.$$

(本书例 8.2.5)

30. 设 $x_i > 0, i = 1, \cdots, n, \sum_{i=1}^{n} x_i = 1$, 则

$$\frac{\sum_{i=1}^{n} x_i}{\prod_{i=1}^{n} x_i} \geqslant n^{n-1}, \quad \sum_{i=1}^{n} \frac{x_i}{1 - x_i} \geqslant \frac{n}{n-1}.$$

等号成立 $\Longleftrightarrow x_i = \frac{1}{n}, i = 1, \cdots, n$.　　(Marshall, Olkin　1979, p. 73)

31. 设 $0 < \min(a_1, a_2, a_3) < c_i \leqslant \max(a_1, a_2, a_3)$, $i = 1, 2, 3$, 且 $\sum_{i=1}^{3} c_i \geqslant \sum_{i=1}^{3} a_i$. 则

(1) $\prod_{i=1}^{3} c_i \geqslant \prod_{i=1}^{3} a_i$;

(2) $\sum_{i \neq j} c_i c_j \geqslant \sum_{i \neq j} a_i a_j$.

(Marshall, Olkin　1979, p. 79)

附录 2　概率统计中的常用不等式

概率统计中常用的不等式主要有两类：一类是矩不等式，它们给出了随机变量不同阶的矩之间的关系；另一类是 Chebyshev 型不等式，它们给出随机变量尾部概率. 本附录主要收集了这两方面的一些重要不等式，另外一部分不等式则是关于经验分布函数及正则化独立和的.

在后面的公式中采用了下列记号. 设 X 为一随机变量

$\mu_r = E(X-\mu)^r$　　r 阶中心矩

$\mu_r' = EX^r$　　r 阶原点矩

$\mu = \mu_1' = EX$　　均值

$\sigma^2 = \mu_2 = E(X-\mu)^2$　　方差

$\nu_r' = E|X|^r$　　r 阶绝对原点矩

$\nu_r = E|X-\mu|^r$　　r 阶绝对中心矩

$\mu_{(r)}' = E[X(X-1)\cdots(X-r+1)]$　　r 阶阶乘矩

$\mu_{(r)} = E[(X-\mu)(X-\mu-1)\cdots(X-\mu-r+1)]$　　r 阶阶乘中心矩

§2.1　矩 不 等 式

1. Cauchy-Schwarz 不等式

$$(E|XY|)^2 \leqslant E|X|^2 E|Y|^2.$$

2. Hölder 不等式

设 $p>1$, $\dfrac{1}{p}+\dfrac{1}{q}=1$, 则

$$E|XY| \leqslant (E|X|^p)^{1/p}(E|Y|^q)^{1/q}.$$

3. C_r－不等式

$$E|X+Y|^r \leqslant C_r(E|X|^r + E|Y|^r), \qquad r>0,$$

其中, 当 $0<r\leqslant 1$, $C_r=1$; 当 $r\geqslant 1$ 时, $C_r = 2^{r-1}$.

4. Minkowski 不等式

$$(E|X+Y|^p)^{1/p} \leqslant (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}, \qquad p>1.$$

(以上诸条见钟开莱　1989, p. 50)

5. $(E|X|^p)^{1/p} \leqslant (E|X|^q)^{1/q}$, $q \geqslant p > 0$. (Loeve 1960, p. 160)

6. $(E|X|^p)^{2p} \leqslant (E|X|^{p-1})^p (E|X|^{p+1})^p$, $p \geqslant 1$. (Rao 1973, p. 149)

7. Liapounov 不等式

$$(E|X|^b)^{a-c} \leqslant (E|X|^c)^{a-b}(E|X|^a)^{b-c}, \ 0 \leqslant c \leqslant b \leqslant a.$$

(Kendall-Stnart 1958, p. 63)

8. Jensen 不等式

设 $g(x)$ 为凸函数, 则

$$g(EX) \leqslant Eg(X).$$ (见本书引理 8.7.1)

9. Gurland 不等式

(a) 设 h_1 和 h_2 为两个单调函数, 其中之一, 譬如 h_1, 是连续函数. 那么当 h_1 和 h_2 同时为增函数或同时为减函数时

$$E(h_1(X)h_2(X)) \geqslant Eh_1(X) \cdot Eh_2(X),$$

当 h_1 和 h_2 有一个是增函数另一个是减函数时, 不等号反向. 特别, 当 X 为非负随机变量时,

$$E(X^{p-1}) \leqslant (EX^p)(EX^{-1}), \qquad p > 0;$$ (Gurland 1967, p. 25)

(b) 设 $h_1, \cdots, h_n$ 为单调连续函数, 且都是增函数或都是减函数, $h_i(x) \geqslant 0$, $i = 1, \cdots, n$. 则

$$E\prod_{i=1}^{n} h_i(x) \geqslant \prod_{i=1}^{n} Eh_i(x).$$

特别, 对任一非负随机变量, 有

$$EX^p \geqslant (EX)^p \geqslant (E(X^{-1}))^{-p} \geqslant (E(X^{-p}))^{-1}, \qquad P \geqslant 1 \text{ 或 } P \leqslant 0.$$

(Gurland 1968, p. 27)

10. Keilson 不等式

(a) 离散分布. 设 $f(x)$ 定义在 $x = 0, 1, 2, \cdots$, 满足 $f^2(x) \geqslant f(x+1)f(x-1)$, 即 $f(x)$ 为对数凹函数, 则

$$\left(\frac{\mu'_{(p+1)}}{(p+1)!}\right)^{1/(p+1)} \leqslant \left(\frac{\mu'_{(p)}}{p!}\right)^{1/p}, \qquad p = 1, 2, \cdots;$$

(b) 连续分布. 设 $f(x)$ 为对数凹函数, 则

$$\left(\frac{\mu'_{p+1}}{(p+1)!}\right)^{1/(p+1)} \leqslant \left(\frac{\mu'_p}{p!}\right)^{1/p} \qquad p=1,2,\cdots;$$

(Keilson 1972, p. 1702~1703)

11. 设 X 为正随机变量

$$E(X^{-p}) \geqslant \frac{E(X^{-(p-1)})}{EX} \geqslant \cdots \geqslant \frac{E(X^{-1})}{(EX)^{p-1}} \geqslant \frac{1}{(EX)^p}, \quad p \geqslant 0;$$

$$E(X^{p-1+q})(EX^p) \leqslant E(X^{p+q})E(X^{p-1}), \quad p>0, q\geqslant 0;$$

(Sclove et al. 1967, p. 34)

12. 设 X 为非负随机变量

$$EX^p \geqslant [E(X^{p/q})]^q \geqslant (EX)^p + [E(X^{p/q}) - (EX)^{p/q}]^q, \quad p\geqslant q\geqslant 1;$$

$$EX^p \geqslant \mu^p + \sigma^p, \quad p\geqslant 2;$$

(Tong 1970, p. 1244)

13. $\mathrm{Cov}(X,Y) \leqslant \mathrm{Var}\left(\dfrac{X+Y}{2}\right)$,

当 X, Y 为正随机变量时, 有

$$\mathrm{Cov}(X,Y) \leqslant \mathrm{Var}\left[\frac{1}{2}(X^\alpha+Y^\alpha)\right]^{1/\alpha}, \quad 0\leqslant\alpha\leqslant 1;$$

$$\mathrm{Cov}\left(X, \frac{Y}{X}\right) \leqslant \mathrm{Var}(Y^{1/2}).$$

(Kimeldorf, Sampson 1973, p. 228, 230)

14. 设 X 和 Y 为独立随机变量, 则

$$E\left(\frac{X}{Y}\right)^p \geqslant \frac{E(X^p)}{E(Y^p)}, \qquad p=1,2,\cdots;$$

$$\mathrm{Var}\left(\frac{X}{Y}\right) \geqslant \frac{\mathrm{Var}(X)}{\mathrm{Var}(Y)}, \quad \text{这里 } EX=EY=0, E(Y^{-1}) \text{ 存在.}$$

(Mullen 1967, p. 30~31)

15. 设 X 为离散型随机变量, $P(X=x_i)=p_i$, $i=1,2,\cdots,n$, 设 $c=\max\limits_i \{x_i\}$, $b=\min\limits_i\{x_i\}$, $a=\sum\limits_{i=1}^n p_i x_i$, 则

$$\mathrm{Var}(X) \leqslant \frac{(c-b)^2}{4};$$

$$\mathrm{Var}(X) \leqslant (c-a)(a-b).$$

(Moors, Muilwijk　1971, p. 385)

16. 设 $X_i(i=1,\cdots,n)$ 为非负随机变量, 则

$$E\left(\sum_{i=1}^{n} X_i\right)^p \geqslant \sum_{i=1}^{n} E(X^p) \text{ 或 } E\left(\sum_{i=1}^{n} X_i\right)^p \leqslant \sum_{i=1}^{n} E(X^p),$$

视 $p \geqslant 1$ 或 $p \leqslant 1$ 而定. 对 $p > 1$, 有

$$E\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right|^p\right] \leqslant \frac{1}{n}\sum_{i=1}^{n} |X_i|^p;$$

$$E\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right|^p\right] \leqslant \left[\frac{1}{n}\sum_{i=1}^{n} (E(|X_i|^p))^{1/p}\right]^p.$$

(钟开莱　1989, p. 52)

17. 设 $X_1,\cdots,X_n$ 为独立随机变量, 且中位数皆为零. 则

$$E\left(\left|\sum_{i=1}^{n} X_i\right|\right) \geqslant \frac{1}{2^{n-1}}\begin{pmatrix} n-1 \\ \left[\frac{n}{2}\right] \end{pmatrix} E\left(\sum_{i=1}^{n} |X_i|\right).$$

这里, $\left[\frac{n}{2}\right]$ 表示小于或等于 $n/2$ 的最大整数.　(Tukey　1946, p. 75)

18. 若 $E_{Y|X}(Y|X) = 0$, 则

$$E|X+Y|^p \geqslant E|X|^p, \quad p \geqslant 1.$$

若 X 和 Y 独立同分布, $EX = 0$, 则

$$\frac{1}{2}E|X-Y|^p \leqslant E|X|^p \leqslant E|X-Y|^p, \quad 1 \leqslant p \leqslant 2.$$

更一般, 对任意的随机变量 X_i $(i=1,\cdots,n)$.

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leqslant \begin{cases} n^{p-1}\left(\sum_{i=1}^{n} E|X_i|^p\right), & p > 1, \\ \sum_{i=1}^{n} E|X_i|^p, & p \leqslant 1. \end{cases}$$

(Patel et al　1976, p. 49)

19. 设 $EX = 0$, $|X| \leqslant M$, 则

$$\mu_p' \leqslant \nu_p' \leqslant M^{p-2}\sigma^2, \quad p \geqslant 2;$$

$$\mu'_p \leqslant M^{p-2}\sigma^2 \left[\frac{1-\left(\dfrac{\sigma}{M}\right)^{2(p-1)}}{1+\left(\dfrac{\sigma}{M}\right)^2}\right] \leqslant M^{p-2}\sigma^2, \quad p=3,5,7,\cdots.$$

(Bennett 1965, p. 560~561)

20. 设 $0<p<1$, ξ_p 为 p 分位点, 则

$$\mu-\sigma\left(\frac{1-p}{p}\right)^{1/2} \leqslant E(X|X\leqslant\xi_p) \leqslant \xi_p \leqslant E(X|X\geqslant\xi_p) \leqslant \mu+\sigma\left(\frac{p}{1-p}\right)^{1/2};$$

$$\xi_q-\xi_p \leqslant E(X|X\geqslant\xi_q)-E(X|X\leqslant\xi_p) \leqslant \sigma\left[\frac{1}{p}+\frac{1}{1-q}\right]^{1/2}, \quad p<q;$$

$$\mu-\frac{1}{2}\nu_1 p \leqslant E(X|X\leqslant\xi_p) \leqslant \xi_p \leqslant E(X|X\geqslant\xi_p) \leqslant \mu+\frac{1}{2}\left(\frac{\nu_1}{1-p}\right);$$

$$|\mu-\xi_{1/2}| \leqslant E(|X-\xi_{1/2}|) \leqslant \nu_1 \leqslant \sigma.$$

(Mallows, Richter 1969, p. 1926~1927)

21. 设 $X, X_1, \cdots, X_n$ 为独立同分布随机变量, $EX=0$. 若对某个正整数 p, $EX^{2p}<\infty$, 则

$$E\left(\sum_{i=1}^n a_i X_i\right)^2 \leqslant EX^2, \quad 若\ p=1;$$

$$E\left(\sum_{i=1}^n a_i X_i\right)^{2p} < \left(\frac{3}{2}\right)^{p-2}(2p-1)!!\,EX^{2p}, \quad 若\ p\geqslant 2,$$

其中 $a_1,\cdots,a_n$ 为满足 $\sum_{i=1}^n a_i^2=1$ 的实数. (陶波, 成平 1981, p. 451~461)

§2.2 Chebyshev 型不等式

1. Chebyshev 不等式

$$P(|X-\mu|\geqslant c\sigma) \leqslant \frac{1}{c^2}, \quad c>0.$$

(陆传荣, 林正炎等 1989, p. 7)

2. Pearson 不等式

$$P(|X-\mu|\geqslant c\nu_p^{1/p}) \leqslant \frac{1}{c^p}, \quad p>0;$$

$$P(|X-\mu| \geqslant c\sigma) \leqslant \frac{v_p}{\sigma^p c^p}, \quad p>0.$$

(Savage　1961, p. 214)

3. Cantelli 不等式

(a)
$$P(|X-\mu| \geqslant k) \leqslant \begin{cases} \dfrac{\nu_p}{c^p}, & \text{若 } c^p \leqslant \dfrac{\nu_{2p}}{\nu_p}, \\ \dfrac{\nu_{2p}-\nu_p^2}{(c^p-\nu_p)^2+\nu_{2p}-\nu_p^2}, & \text{若 } c^p \geqslant \dfrac{\nu_{2p}}{\nu_p}; \end{cases}$$

(Sarage　1961, p. 217)

(b)
$$P(X-\mu \leqslant c) \leqslant \frac{\sigma^2}{\sigma^2+c^2}, \quad \text{若 } c<0,$$

$$P(x-\mu \leqslant c) \geqslant 1-\frac{\sigma^2}{\sigma^2+c^2}, \quad \text{若 } c \geqslant 0.$$

(Rao　1973, p. 145)

4. Peek 不等式

$$P(|X-\mu| \geqslant c\sigma^2) \leqslant \frac{1-\delta^2}{c^2-2c\delta+1}, \qquad \text{若 } c \geqslant \delta.$$

(Rao　1973, p. 145)

这里 $\delta = \nu_1/\sigma$.

对于众数等于均值 μ 的单峰分布, 有

$$P(|X-\mu| \geqslant c\sigma) \leqslant \frac{4}{9}\frac{1-\delta^2}{(c-\delta)^2}, \quad \text{若} c \geqslant \delta.$$

(Savage　1961, p. 217)

5. Markov 不等式

$$P(|X| \geqslant c) \leqslant \frac{v_{p'}}{c^p}, \quad \text{若} c>0.$$

(Loeve　1960, p. 158)

6. Camp-Meidell 不等式

设 X 具有单峰分布, m_0 为其众数, 记 $\tau^2 = \sigma^2 + (\mu - m_0)^2$. 则

$$P(|X-m_0| \geqslant c\tau) \leqslant \begin{cases} 1-\dfrac{c}{\sqrt{3}}, & \text{若} c \leqslant \dfrac{2}{\sqrt{3}}, \\ \dfrac{4}{9c^2}, & \text{若} c \geqslant \dfrac{2}{\sqrt{3}}; \end{cases}$$

$$P(|X-\mu| \geqslant c\sigma) \leqslant \frac{4}{9}\frac{(1+s^2)}{(c-s)^2},$$

(Savage　1961, p. 216)

其中 $s = |\mu - m_0|/\sigma, c > s$.

7. Selberg 不等式

设 $-\alpha < 0 < \beta, m = \min(\alpha, \beta)$, 则

$$P(-\alpha < X - \mu < \beta) \geqslant \begin{cases} \dfrac{4(\alpha\beta - \sigma^2)}{(\alpha+\beta)^2}, & \text{若}\alpha\beta - m^2 \leqslant 2\sigma^2 \leqslant 2\alpha\beta, \\ \dfrac{m^2}{\alpha^2 + m^2}, & \text{若}2\alpha^2 \leqslant \alpha\beta - m^2, \\ 0, & \text{若}\alpha\beta \leqslant m^2. \end{cases}$$

(Karlin, Studden 1966, p. 475)

8. Glasser 不等式

设 $\frac{1}{2}\left(\frac{1}{c_1} + \frac{1}{c_2}\right) \leqslant 1$, 则

$$P(-c_1 v_1 < X - \mu < c_2 v_1) \geqslant 1 - \frac{1}{2}\left(\frac{1}{c_1} + \frac{1}{c_2}\right).$$

(Karlin, Studden 1966, p. 481)

9. 设 $g(x) \geqslant 0$, 则对 $c > 0$, 有

$$P(g(X) \geqslant c) \leqslant \frac{Eg(X)}{c}.$$

(Hogg, Craig 1970, p. 54)

10. 设 $g(x) \geqslant d > 0$, 则

$$P(X \geqslant c) \leqslant \frac{Eg(X)}{d}.$$

(Fisz 1963, p. 101)

11. 设 $g(x)$ 为正值偶函数, 当 $x > 0$ 时, 为非减函数. 则

$$P(|X| \geqslant c) \leqslant \frac{Eg(X)}{g(c)}, \quad c > 0.$$

(钟开莱 1989, p. 51)

12. 设 $g(x)$ 为非负偶函数, 在 $x > 0$ 上为非减函数. 对一切 $x, g(x) \leqslant M$. 则

$$P(|X| \geqslant M) \geqslant \frac{Eg(X) - g(M)}{M}.$$

(Fisz 1963, p. 102)

13. 设 $g(x)$ 为正值偶函数, 在 $x>0$ 上为非减数. 设 X 满足 $P(|X|\leqslant M)=1$, 则对一切 $c>0$.

$$P(|X|\geqslant c)\geqslant\frac{Eg(X)-g(c)}{g(M)}.$$

(Fisz 1963, p. 102)

14. 设对 $\alpha>0, E(e^{\alpha x})$ 存在, 则

$$P(X\geqslant c)\leqslant\frac{Ee^{\alpha x}}{e^{\alpha c}}.$$

(Rao 1973, p. 143)

15. 设 $g(x)$ 为正的非减函数, 则

$$P(|X-\mu|\geqslant c)\leqslant\frac{E(g(|X-\mu|))}{g(c)}.$$

(Gnedenko 1962, p. 249)

16. 设 $EX=0$, 则对 $c>1$, 有

$$P(|X|\geqslant c\sigma)\geqslant\frac{\mu_4-\sigma^4}{\mu_4+c^4\sigma^4-2c^2\sigma^4}.$$

(陆传荣等 1989, p. 46)

17. Kolmogorov 不等式

(a) 设 $X_i(i=1,\cdots,n)$ 为独立随机变量. $\mathrm{Var}(X_i)=\sigma_i^2<\infty, s_n=\sum\limits_{i=1}^{n}X_i, \sigma^2=\sum\limits_{i=1}^{n}\sigma_i^2$. 则对任意 $\varepsilon>0$,

$$P\{\max_{1\leqslant k\leqslant n}|S_k-ES_k|\geqslant\varepsilon\}\leqslant\frac{\sigma^2}{\varepsilon^2};$$

(b) 设 $X_i(i=1,\cdots,n)$ 为独立随机变量, 假设存在 $c>0, |X_i|\leqslant c, i=1,\cdots,n.\sigma^2$ 的定义同 (a), 则对任意 $\varepsilon>0$,

$$P\{\max_{1\leqslant k\leqslant n}|S_k-E(S_k)|\geqslant\varepsilon\}\geqslant 1-\frac{(\varepsilon+c)^2}{\sigma^2}.$$

(陆传荣等 1989, p. 139)

18. Hajek-Renyi 不等式

设 $X_i(i=1,\cdots,n)$ 为独立随机变量, $EX_i=0, \mathrm{Var}(X_i)=\sigma_i^2<\infty, c_1, c_2, \cdots$ 为正数非增序列, $S_n=\sum\limits_{i=1}^{n}X_i$. 则对任意正整数 $m, n(m<n)$ 和 $\varepsilon>0$, 有

$$P\{\max_{m\leqslant k\leqslant n}c_k|S_k|\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon^2}\left(c_m^2\sum_{i=1}^{m}\sigma_i^2+\sum_{i=m+1}^{n}c_i^2\sigma_i^2\right).$$

(Rao 1973, p. 143)

19. Heyde 不等式

设 $X_1, \cdots, X_n$ 为 n 个随机变量, $s_n = \sum_{i=1}^n X_i, Y_1, \cdots, Y_n$ 为 n 个独立随机变量, $EY_i = 0, \text{Var}(Y_i) < \infty, i = 1, \cdots, n$, 记 $Z_i = X_i - Y_i$, 设 $c_1, c_2, \cdots$ 为正数的非增序列, 则对任意正整数 $m, n(m < n), \varepsilon > 0, 0 < \alpha < 1$, 有

$$P\{\max_{m \leqslant k \leqslant n} c_k|s_k| \geqslant \varepsilon\} \leqslant \frac{c_m^2\left(\sum_{i=1}^m EY_i^2\right) + \sum_{m+1}^n c_i^2(Y_i^2)}{(1-\alpha)^2\varepsilon^2} + \sum_{i=m+1}^n P(Z_i \neq 0) + P\left(c_n\left|\sum_{i=1}^n Z_i\right| \geqslant n\varepsilon\right).$$

(Tomkins 1971, p. 428)

20. Kounias-Weng 不等式

(a) 设 $X_i (i = 1, \cdots, n)$ 为 n 个随机变量, $E|X_i|^p < \infty, i = 1, \cdots, n$, 这里 $0 < p \leqslant 1$. 记 $S_n = \sum_{i=1}^n X_i$. 设 $c_1, c_2, \cdots$ 为正整数非增序列. 则对任意正整数 $m, n(m < n), \varepsilon > 0$, 有

$$P\{\max_{m \leqslant k \leqslant n} c_k|S_k| \geqslant \varepsilon\} \leqslant \left(c_p^m \sum_{i=1}^m E|X_i|^p + \sum_{i=m+1}^n c_i^p E|X_i|^p\right) / \varepsilon^p;$$

(b) 在上款假设下, 若 $p \geqslant 1$, 则

$$P\{\max_{m \leqslant k \leqslant n} c_k|S_k| \geqslant \varepsilon\} \leqslant \left\{c_m \sum_{i=1}^m (E|X_i|^p)^{1/p} + \sum_{m+1}^n c_i (E|X_i|^p)^{1/p}\right\}^p / \varepsilon^p.$$

(Kounias, Weng 1969, p. 1091~1092)

21. Markov 不等式

设 $X_1, \cdots, X_n$ 为非负随机变量, $S_n = \sum_{i=1}^n X_i, \mu_i = EX_i, i = 1, \cdots, n$, 则对任给 $\varepsilon > 0$,

$$P\left\{S_n \geqslant \varepsilon \sum_{i=1}^n \mu_i\right\} \leqslant \frac{1}{\varepsilon}.$$

(Savage 1961, p. 213)

22. 设 $X_1, \cdots, X_n$ 为 n 个独立随机变量, $P(|X_i + \cdots + X_n| > \varepsilon) \leqslant \alpha, i = 1, \cdots, n$. 记 $S_n = \sum_{i=1}^n X_i$. 则

$$P\{\max_{1 \leqslant k \leqslant n} |S_k| > 2\varepsilon\} \leqslant P(|S_n| > \varepsilon)/(1-\alpha).$$

(Rao 1973, p. 145)

23. Berge 不等式

设 X_1, X_2 为两个随机变量, $EX_i = \mu_i, \mathrm{Var}(X_i) = \sigma_i^2$, $\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$. 则

$$P\left\{\max\left(\frac{|X_1-\mu_1|}{\sigma_1}, \frac{|X_2-\mu_2|}{\sigma^2}\right) \geqslant \varepsilon\right\} \leqslant \frac{1+\sqrt{1-\rho^2}}{\varepsilon^2}.$$

(Rao　1973, p. 145)

24. Birnbaum-Raymond-Zuckerman 不等式

设 $X_1, \cdots, X_n$ 为 n 个独立随机变量, $\mu_i = EX_i, \sigma^2 = \mathrm{Var}(X_i), i = 1, \cdots, n$, 若 n 为偶数, 则对任给 $\varepsilon > 0$,

$$P\left\{\sum_{i=1}^{n}(X_i-\mu_i)^2 \geqslant \varepsilon^2\right\} \leqslant \begin{cases} 1, & \text{若 } \varepsilon^2 \leqslant n\sigma^2, \\ \dfrac{n\sigma^2}{2\varepsilon^2 - n\sigma^2}, & \text{若 } n\sigma^2 \leqslant \varepsilon^2 \leqslant \dfrac{n\sigma^2}{4}(3+\sqrt{5}), \\ \dfrac{n\sigma^2}{\varepsilon^2}\left(1-\dfrac{n\sigma^2}{4\varepsilon^2}\right), & \text{若 } \dfrac{n\sigma^2}{4}(3+\sqrt{5}) \leqslant \varepsilon^2; \end{cases}$$

若 n 为奇数

$$P\left(\sum_{i=1}^{n}(X_i-\mu_i)^2 \geqslant \varepsilon^2\right) \leqslant \begin{cases} 1, & \text{若 } \varepsilon^2 \leqslant n\sigma^2, \\ \dfrac{(n+1)\sigma^2}{2\varepsilon^2-(n-1)\sigma^2}, & \text{若 } n\sigma^2 \leqslant \varepsilon^2 \leqslant \dfrac{\sigma^2}{4}d_n, \\ \dfrac{n\sigma^2}{\varepsilon^2} - \dfrac{(n^2-1)}{4}\dfrac{\sigma^4}{\varepsilon^4}, & \text{若 } \varepsilon^2 \geqslant \dfrac{\sigma^2}{4}d_n, \end{cases}$$

其中

$$d_n = 3n + 1 + (5n^2 + 6n + 5)^{1/2}.$$

(Savage　1961, p. 214)

25. Guttman 不等式

设 $X_1, \cdots, X_2$ 为 n 个独立同分布随机变量, $\mu = EX_i, \sigma^2 = \mathrm{Var}(X_i), i = 1, \cdots, n$. 记

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i, s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

则对 $c \geqslant 1$, 有

$$P\left\{(\overline{X}-\mu)^2 \geqslant \frac{s^2}{n-1} + \sigma^2\left[\frac{2(\varepsilon^2-1)}{n(n-1)}\right]^{1/2}\right\} \leqslant \frac{1}{c^2}.$$

(Savage　1961, p. 215)

26. Bernstein 不等式

设 $X_1, \cdots, X_n$ 为 n 个独立随机变量, $\mu_i = EX_i, \sigma_i^2 = \mathrm{Var}(X_i)$, $i = 1, \cdots, n$. $P(|X_i - \mu_i| > m) = 0$, 这里 $m < \infty$. 记

$$\mu = \sum_{i=1}^{n} \mu_i, \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2, \quad s_n = \sum_{i=1}^{n} X_i.$$

则对任意 $\varepsilon > 0$,

$$P\{|s_n - \mu| \geqslant n\varepsilon\} \leqslant 2\exp\left\{\frac{-n^2\varepsilon^2}{2\sigma^2 + \dfrac{2}{3}mn\varepsilon}\right\}.$$

(Serfling 1980, p. 95)

27. Hoeffding 不等式

(a) 设 $X_1, \cdots, X_n$ 为 n 个独立随机变量, $0 \leqslant X_i \leqslant 1$, $i = 1, \cdots, n$. $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i, \mu = \dfrac{1}{n}\sum_{i=1}^{n} EX_i$, 则对 $\varepsilon > 0$,

$$\begin{aligned} P\{\overline{X} - \mu \geqslant \varepsilon\} &\leqslant \left[\left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon}\left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon}\right]^n \\ &\leqslant \exp\{-n\varepsilon^2 g(\mu)\} \\ &\leqslant \exp\{-2n\varepsilon^2\}, \end{aligned}$$

其中

$$g(\mu) = \begin{cases} \dfrac{1}{1-2\mu}\log\left(\dfrac{1-\mu}{\mu}\right), & \text{若 } 0 < \mu < \dfrac{1}{2}, \\ \dfrac{1}{2\mu(1-\mu)}, & \text{若 } \dfrac{1}{2} \leqslant \mu < 1. \end{cases}$$

(b) 设 $X_1, \cdots, X_n$ 为独立随机变量, $a_i \leqslant X_i \leqslant b_i$, $i = 1, \cdots, n$. 则对任意 $\varepsilon > 0$,

$$P(\overline{X} - \mu \geqslant \varepsilon) \leqslant \exp\left\{\frac{-2n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\}.$$

(c) 设 $X_1, \cdots, X_n$ 为独立随机变量, $EX_i = 0, X_i \leqslant b, i = 1, \cdots, n$. 则对 $0 < \varepsilon < b$, 有

$$P(\overline{X} \geqslant \varepsilon) \leqslant \exp\left\{-\frac{n\varepsilon}{b}\left[\left(1 + \frac{\sigma^2}{b\varepsilon}\right)\log\left(1 + \frac{b\varepsilon}{\sigma^2}\right) - 1\right]\right\}.$$

(Karlin, Studden 1966, p. 535~536)

28. Bennett 不等式

设 $X_1,\cdots,X_n$ 为独立随机变量, $EX_i=0,\mathrm{Var}(X_i)=\sigma_i^2,E|X_i|^2\leqslant m^{p-2}\sigma_i^2,p=2,3\cdots$, 其中 m 为常数, 记 $n\sigma^2=\sum_{i=1}^n\sigma_i^2$. 则

$$P(\overline{X}\geqslant\varepsilon)\leqslant\exp\left\{-\frac{n\varepsilon}{m}\left[\left(1+\frac{\sigma^2}{m\varepsilon}\right)\log\left(1+\frac{m\varepsilon}{\sigma^2}\right)-1\right]\right\}.$$

(Karlin, Studden 1966, p. 537)

§2.3 其 他

1. 关于 Kolmogorov-Smirnov 距离的不等式

设 $X_1,\cdots,X_n$ 为来自总体 $F(x)$ 的简单随机样本. 经验分布函数 $F_n(x)$ 定义为

$$F_n(x)=\frac{1}{n}\sum_{i=1}^n I(X_i\leqslant x),$$

其中 $I(A)$ 表示随机事件 A 的示性函数, $F_n(x)$ 与 $F(x)$ 接近程度的度量

$$D_n=\sup_{-\infty<x<\infty}|F_n(x)-F(x)|$$

称为 Kolmogorov-Smirnov 距离.

(a) Dvoretzky-Kiefer-Wolfowitz 不等式

对任意 $\varepsilon>0$, 存在一个不依赖于总体分布 F 的有限正常数 c, 使得

$$P(D_n>\varepsilon)\leqslant ce^{-2n\varepsilon^2}, n=1,2,\cdots;$$

(b) 对任意 $\varepsilon>0$,

$$P\left(\sup_{m\geqslant n}D_m>\varepsilon\right)\leqslant\frac{c}{1-\rho_\varepsilon}\rho_\varepsilon^n,$$

其中 $\rho_\varepsilon=\exp(-2\varepsilon^2)$, c 的定义同 (a) 款. (Serfling 1980, p. 59~60)

2. Berry – Esseen 不等式

设 $X_1,\cdots,X_n$ 为独立同分布随机变量, 定义正则化和 S_n^* 为

$$S_n^*=\frac{\sum_{i=1}^n X_i-E\left(\sum_{i=1}^n X_i\right)}{\left(\mathrm{Var}\sum_{i=1}^n X_i\right)^{1/2}}.$$

其分布函数记为 $G_n(t)$, 即 $G_n(t) = P(S_n^* \leqslant t)$.

若 $EX_i = \mu, \mathrm{Var}(X_i) = \sigma^2$, 则 $S_n^* = \sqrt{n}(\overline{X} - \mu)/\sigma$. 那么

$$\sup_t |G_n(t) - \varPhi(t)| \leqslant \frac{33}{4} \frac{E|X_1 - \mu|^3}{\sigma^3 n^{1/2}}, \quad n = 1, 2, \cdots.$$

其中 $\varPhi(x)$ 为 $N(0,1)$ 的分布函数.

注 1 许多作者试图改进右边的常数 $\frac{33}{4}$, Zolotarev 把此数降为 0.91, 而 Beeck 进一步减为 0.7975.

注 2 若 $X_1, \cdots, X_n$ 只是独立, 而不同分布, 记 $\mu_i = EX_i$ 则上面的不等式右端改为

$$c \frac{\sum_{i=1}^{n} E|X_i - \mu_i|^3}{\left[\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)\right]^{3/2}}.$$

(Serfling 1980, p.33)